

(51)Int.Cl.[®]
H 0 4 N 7/32

識別記号

F I
H 0 4 N 7/137

Z

審査請求 未請求 請求項の数27 OL 外国語出願 (全 79 頁)

(21)出願番号 特願平10-98040
 (22)出願日 平成10年(1998)3月9日
 (31)優先権主張番号 4 0 1 2 0
 (32)優先日 1997年3月7日
 (33)優先権主張国 米国(US)
 (31)優先権主張番号 4 2 2 4 5
 (32)優先日 1997年3月31日
 (33)優先権主張国 米国(US)
 (31)優先権主張番号 9 4 4 1 1 8
 (32)優先日 1997年10月6日
 (33)優先権主張国 米国(US)

(71)出願人 597139435
 ネクストレベル・システムズ・インコーポ
 レイテッド
 アメリカ合衆国イリノイ州シカゴ、ウエス
 トプリンマウアー・アベニュー8770, サー
 ティーンス・フロア
 (72)発明者 ロパート・オー・エイフリグ
 アメリカ合衆国カリフォルニア州サンディ
 エゴ、カミニト アルバレス 10530
 (74)代理人 弁理士 竹内 澄夫 (外1名)

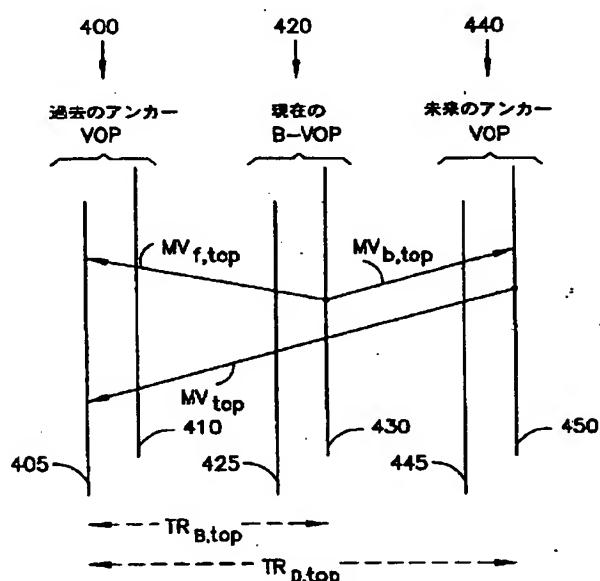
最終頁に続く

(54)【発明の名称】インターレースされたデジタルビデオ用の双方向予測ビデオオブジェクト平面の予測及び符号化

(57)【要約】 (修正有)

【課題】B-VOP内のMBに対し、移動ベクトル(MV)予測を与える効果的な方法およびデコーダを提供する。

【解決手段】未来のアンカー画像440のフィールド予測したマクロブロックとともに、共同配置されたB-VOPマイクロブロック420に対して、直接モードの予測が四つのフィールド移動ベクトル($MV_{f,top}$, $MV_{f,bot}$, $MV_{b,top}$, $MV_{b,bot}$)を計算することによりなされる。四つのフィールド移動ベクトルおよびそれらの基準フィールドは(1)現在のマクロブロックのコード化ベクトルのオフセット期間(MVD), (2)二つの未来のアンカー画像フィールド移動ベクトル(MV_{top} , MV_{bot}), (3)共同配置される未来のアンカーマクロブロックの二つのフィールド移動ベクトルにより使用される基準フィールド405, 410, および(4)現在のB-VOPフィールドとアンカーフィールドとの間、フィールド期間における時間間隔($TR_{B,top}$, $TR_{b,top}$, $TR_{D,top}$, $TR_{D,bot}$)から決定される。



1

【特許請求の範囲】

【請求項1】 一連のデジタルビデオイメージにおける、トップおよびボトムフィールドを有する、現在の、双方向に推測した、フィールドコード化イメージに対する、直接モード移動ベクトルを計算する方法であって、トップおよびボトムフィールドを有する過去のフィールドコード化した基準イメージ、およびトップおよびボトムフィールドを有する未来のフィールドコード化基準のイメージを決定する工程であって、ここで、未来のイメージが未来のイメージのトップフィールドのフォワード移動ベクトル、 MV_{top} が過去のイメージのトップおよびボトムフィールドの一つを基準とし、未来のイメージのボトムフィールドのフォワード移動ベクトル、 MV_{bot} が過去のイメージのトップおよびボトムフィールドの一つを基準とするように、過去のイメージを使用して予測される、ところの工程と、未来のイメージの、対応するフィールドのフォワード移動ベクトルをスケーリングすることにより、現在のイメージのトップおよびボトムフィールドの少なくとも一つを予測するための、フォワードおよびバックワード移動ベクトルを決定する工程と、を含む方法。

【請求項2】 請求項1に記載の方法であって、現在のイメージのトップフィールドを予測するためのフォワード移動ベクトル、 $MV_{f,top}$ は式、 $(MV_{top} * TR_{B,top}) / TR_{D,top} + MV_D$ にしたがって決定され、ここで、 $TR_{B,top}$ は現在のイメージのトップフィールドと、 MV_{top} により基準とされる過去のイメージのフィールドとの間の時間的な間隔に対応し、 $TR_{D,top}$ は未来のイメージのトップフィールドと、 MV_{top} により基準とされる過去のイメージのフィールドとの間の時間的な間隔に対応し、 MV_D はデルタ移動ベクトルである、ところの方法。

【請求項3】 請求項2に記載の方法であって、 $MV_{f,top}$ は、ゼロの方向へのトラケーションをもつ整数の分数を使用して決定され、 MV_{top} および MV_{bot} は整数の1/2ルマペル移動ベクトルである、ところの方法。

【請求項4】 請求項2または3に記載の方法であって、 $TR_{B,top}$ および $TR_{D,top}$ は、前記現在のフィールドコード化したイメージが最初にトップフィールドであるか最初にボトムフィールドであるかどうかを説明する時間補正を組み入れる、ところの方法。

【請求項5】 請求項1ないし4のいずれかに記載の方法であって、現在のイメージのボトムフィールドを予測するためのフォワード移動ベクトル、 $MV_{f,bot}$ は、式 $(MV_{bot} * TR_{B,top}) / TR_{D,bot} + MV_D$ にしたがって、決定され、ここで、 $TR_{B,bot}$ は現在のイメージのボトムフィールドと、 MV_{bot} により基準とされる過去のイメージのフィールドとの間の時間的な間隔に対応し、 $TR_{D,bot}$ は未来のイメージのボトムフィールドと、 MV_{bot} により基準とされる過去のイメージのフィールドとの間の時間的な間隔に対応し、 MV_D はデルタ移動ベクトルである、ところの

2

方法。

【請求項6】 請求項5に記載の方法であって、 $MV_{f,top}$ は、ゼロの方向へのトラケーションをもつ整数の分数を使用して決定され、 MV_{top} および MV_{bot} は整数の1/2ルマペル移動ベクトルである、ところの方法。

【請求項7】 請求項5または6に記載の方法であって、 $TR_{B,bot}$ および $TR_{D,tbot}$ は、前記現在のフィールドコード化したイメージが最初にトップフィールドであるか最初にボトムフィールドであるかどうかを説明する時間補正を組み入れる、ところの方法。

【請求項8】 請求項1ないし7のいずれかに記載の方法であって、現在のイメージのトップフィールドを予測するためのバックワード移動ベクトル、 $MV_{B,top}$ は式

(a) $MV_{B,top} = ((TR_{B,top} - TR_{D,top}) * MV_{top}) / T$
 RD,top 、および (b) $MV_{B,top} = MV_{f,top} - MV_{top}$ の一方にしたがって、決定され、ここで、 $TR_{B,top}$ は現在のイメージのトップフィールドと、 MV_{top} により基準とされる過去のイメージのフィールドとの間の時間的な間隔に対応し、 $TR_{D,top}$ は未来のイメージのトップフィールドと、 MV_{top} により基準とされる過去のイメージのフィールドとの間の時間的な間隔に対応し、 $MV_{f,top}$ は現在のイメージのトップフィールドを予測するためのフォワード移動ベクトルである、ところの方法。

【請求項9】 請求項8に記載の方法であって、前記式(a)は、デルタ移動ベクトル、 $MV_D=0$ のときを選択され、前記式(b)は、 $MV_D \neq 0$ のとき選択される、ところの方法。

【請求項10】 請求項1ないし9のいずれかに記載の方法であって、現在のイメージのボトムフィールドを予測するためのバックワード移動ベクトル、 $MV_{b,bot}$ は、

(a) $MV_{b,bot} = ((TR_{B,bot} - TR_{D,bot}) * MV_{bot}) / T$
 RD,bot 、および (b) $MV_{b,bot} = MV_{f,bot} - MV_{bot}$ の一方にしたがって、決定され、ここで、 $TR_{B,bot}$ は現在のイメージのボトムフィールドと、 MV_{bot} により基準とされる過去のイメージのフィールドとの間の時間的な間隔に対応し、 $TR_{D,bot}$ は未来のイメージのボトムフィールドと、 MV_{bot} により基準とされる過去のイメージのフィールドとの間の時間的な間隔に対応し、 $MV_{f,bot}$ は現在のイメージのボトムフィールドを予測するためのフォワード移動ベクトルである、ところの方法。

【請求項11】 請求項10に記載の方法であって、前記式(a)は、デルタ移動ベクトル、 $MV_D=0$ のときを選択され、前記式(b)は、 $MV_D \neq 0$ のとき選択される、ところの方法。

【請求項12】 一連のデジタルビデオイメージにおいて、トップおよびボトムのフィールドを有する、現在の、予測された、フィールドコード化マクロブロックのためのコード化モードを選択する方法であって、過去の基準マクロブロック(フォワードコード化モードに対応する)に関し、現在のマクロブロックに対する絶対差分

50

エラーのフォワード合計, SAD_{forward, field}を決定する工程と, 未来の基準マクロブロック(バックワードコード化モードに対応する)に関し, 現在のマクロブロックに対する絶対差分エラーのフォワード合計, SAD_{backward, field}を決定する工程と, 未来および過去の基準マクロブロック(平均コード化モードに対応する)に関し, 現在のマクロブロックに対する絶対差分エラーの平均合計, SAD_{average, field}を決定する工程と, 前記SADの最小にしたがって前記コード化モードを選択する工程と, を含む方法。

【請求項13】 請求項12に記載の方法であって, それぞれのコード化モードの必要な移動ベクトルの数を説明する対応するバイアス期間とともに, 前記SADのそれぞれの合計の最小にしたがって前記コード化モードを選択する工程を, さらに含むところの方法。

【請求項14】 請求項12または13に記載の方法であって, SAD_{forward, field}は, (a) 過去の基準のマイクロブロックのトップフィールドに関し, 現在のマクロブロックのトップフィールドに対する絶対的な差分の合計, および(b) 過去の基準のマイクロブロックのボトムフィールドに関し, 現在のマクロブロックのボトムフィールドに対する絶対的な差分の合計の総計にしたがって決定される, ところの方法。

【請求項15】 請求項12または13に記載の方法であって, SAD_{backward, field}は, (a) 未来の基準のマイクロブロックのトップフィールドに関し, 現在のマクロブロックのトップフィールドに対する絶対的な差分の合計, および(b) 未来の基準のマイクロブロックのボトムフィールドに関し, 現在のマクロブロックのボトムフィールドに対する絶対的な差分の合計の総計にしたがって決定される, ところの方法。

【請求項16】 請求項12ないし15のいずれかに記載の方法であって, SAD_{average, field}は, (a) 過去および未来の基準のマイクロブロックのトップフィールドの平均に関し, 現在のマクロブロックのトップフィールドに対する絶対的な差分の合計, および(b) 過去および未来の基準のマイクロブロックのボトムフィールドの平均に関し, 現在のマクロブロックのボトムフィールドに対する絶対的な差分の合計の総計にしたがって決定される, ところの方法。

【請求項17】 受信したビットストリームからの一連のデジタルビデオイメージにおける, トップおよびボトムのフィールドを有する, 現在の, 直接モードのフィールドコード化したマクロブロックであって, トップおよびボトムのフィールドを有する過去のフィールドコード化した基準マクロブロック, およびボトムフィールドを有する未来の, フィールドコード化した基準マクロブロックを使用して双方向に予測されるところのマクロブロックを回復するデコーダであって, 過去のマクロブロックのトップおよびボトムのフィールドの一つを基準とす

る未来のマクロブロックのトップフィールドのフォワード移動ベクトル, MV_{top}, ならびに前記過去のトップおよびボトムのフィールドの一つを基準とする未来のマクロブロックのボトムフィールドのフォワード移動ベクトル, MV_{bot}を回復する手段と, 未来のマクロブロックの, 対応するフィールドのフォワード移動ベクトルをスケーリングすることにより, 現在のマクロブロックのトップおよびボトムのフィールドの少なくとも一つを予測するため, フォワードおよびバックワード移動ベクトルを決定する手段と, を含むデコーダ。

【請求項18】 請求項17に記載のデコーダであって, さらに式 $(MV_{top} * TR_{B, top}) / TR_{D, top} + MV_D$ にしたがって現在のマクロブロックのトップフィールドを予測するため, フォワード移動ベクトル, MV_{f, top}を決定する手段を含み, ここで, MV_{B, top}は現在のマクロブロックのトップフィールドと, TR_{top}により基準とされる過去のイメージのフィールドとの間の時間的な間隔に対応し, MV_{D, top}は未来のイメージのトップフィールドと, MV_{top}により基準とされる過去のマクロブロックのフィールドとの間の時間的な間隔に対応し, MV_Dはデルタ移動ベクトルである, ところのデコーダ。

【請求項19】 請求項18に記載のデコーダであって, MV_{f, top}は, ゼロの方向へのトラケーションをもつ整数の分数を使用して決定され, MV_{top}およびMV_{bot}は整数の1/2ルマペル移動ベクトルである, ところのデコーダ。

【請求項20】 請求項18または19に記載のデコーダであって, TR_{B, top}およびTR_{D, top}は, 前記現在のフィールドコード化したマクロブロックが最初にトップフィールドまたは最初にボトムフィールドであるかどうかを説明する時間補正を組み入れる, ところのデコーダ。

【請求項21】 請求項17ないし20のいずれかに記載のデコーダであって, 式 $(MV_{bot} * TR_{B, top}) / TR_{D, bot} + MV_D$ にしたがって, 現在のマクロブロックのボトムフィールドを予測するため, フォワード移動ベクトル, MV_{f, bot}を決定する手段を含み, ここで, TR_{B, bot}は現在のマクロブロックのボトムフィールドと, MV_{bot}により基準とされる過去のマクロブロックのフィールドとの間の時間的な間隔に対応し, TR_{D, bot}は未来のマクロブロックのボトムフィールドと, MV_{bot}により基準とされる過去のマクロブロックのフィールドとの間の時間的な間隔に対応し, MV_Dはデルタ移動ベクトルである, ところのデコーダ。

【請求項22】 請求項21に記載のデコーダであって, MV_{f, top}は, ゼロの方向へのトラケーションをもつ整数の分数を使用して決定され, MV_{top}およびMV_{bot}は整数の1/2ルマペル移動ベクトルである, ところのデコーダ。

【請求項23】 請求項21または22に記載のデコーダであって, TR_{B, bot}およびTR_{D, bot}は, 前記現在のフィールドコード化したマクロブロックが最初にトップフィー

ルドまたは最初にボトムフィールドであるかどうかを説明する時間補正を組み入れる、ところのデコーダ。

【請求項24】 請求項17ないし23のいずれかに記載のデコーダであって、式(a) $MV_{b,top} = ((TR_{B,top}-T_{RD,top}) * MV_{top}) / TR_{D,top}$ 、および(b) $MV_{b,top}=MV_{f,top}-MV_{top}$ の一方にしたがって、現在のマクロブロックのトップフィールドを予測するため、バックワード移動ベクトル、 $MV_{B,top}$ を決定する手段を含み、ここで、 $T_{RB,top}$ は現在のマクロブロックのトップフィールドと、 MV_{top} により基準とされる過去のマクロブロックのフィールドとの間の時間的な間隔に対応し、 $TR_{D,top}$ は未来的マクロブロックのトップフィールドと、 MV_{top} により基準とされる過去のマクロブロックのフィールドとの間の時間的な間隔に対応し、 $MV_{f,top}$ は現在のマクロブロックのトップフィールドを予測するためのフォワード移動ベクトルである、ところのデコーダ。

【請求項25】 請求項24に記載のデコーダであって、さらにデルタ移動ベクトル $MV_D=0$ のとき前記式(a)を選択する手段と、 $MV_D \neq 0$ のとき前記式(b)を選択する手段と、含むところのデコーダ。

【請求項26】 請求項1ないし9のいずれかに記載のデコーダであって、(a) $MV_{b,bot} = ((TR_{B,bot}-T_{RD,bot}) * MV_{bot}) / TR_{D,bot}$ 、および(b) $MV_{b,bot}=MV_{f,bot}-MV_{bot}$ の一方にしたがって、現在のマクロブロックのボトムフィールドを予測するため、バックワード移動ベクトル、 $MV_{B,bot}$ を決定する手段を含み、ここで、 $T_{RB,bot}$ は現在のマクロブロックのボトムフィールドと、 MV_{bot} により基準とされる過去のマクロブロックのフィールドとの間の時間的な間隔に対応し、 $TR_{D,bot}$ は未来的マクロブロックのボトムフィールドと、 MV_{bot} により基準とされる過去のマクロブロックのフィールドとの間の時間的な間隔に対応し、 $MV_{f,bot}$ は現在のマクロブロックのボトムフィールドを予測するためのフォワード移動ベクトルである、ところのデコーダ。

【請求項27】 請求項26に記載のデコーダであって、さらにデルタ移動ベクトル、 $MV_D=0$ のときに前記式(a)を選択する手段、および $MV_D \neq 0$ のとき前記式(b)を選択する手段を含む、ところのデコーダ。

【発明の詳細な説明】

【0001】

【発明の属する技術分野】 本発明は双方向に予測されたビデオオブジェクトプレーン(B-BOP)のようなデジタルビデオイメージ(特に、B-VOPおよび/またはB-VOPをコード化するために使用される基準イメージがインターレース・コード化される場合)をコード化するための方法および装置に関する。

【0002】

【従来の技術】 本発明は特にさまざまなマルチメディア応用とともに使用するのが好適であり、ここに参考文献として組み込む文献ISO/IEC/JTC1/SC29/WG11 N1796内の

題名“MPEG-4 Video Verification Model Version 8.0”(ストックホルム、1997年7月)に記載されている、MPEG Verification Model(VM) 8.0標準規格(MPEG-4 VM 8.0)と互換性がある。MPEG-2標準規格はMPEG-4標準規格に先行するものであり、ここに参考文献として組み込む文献ISO/IEC 13818-2題名“Information Technology- Generic Coding of Moving Pictures and Associated Audio, Recommendation H.262”1994年3月25日に記載されている。

【0003】 MPEG-4は、柔軟なフレームワーク、並びにデジタルオーディオ-ビジュアルデータの通信、アクセス及び操作のためのコード化ツールのオープンセットを与える新しい符号化標準規格である。これらのツールは広範囲な特徴を支援する。MPEG-4の柔軟なフレームワークは、コンピュータ、テレコミュニケーション、およびエンターテイメント(すなわち、TV及び映画)産業に要求されるデータベースキャッシング、情報検索、および双方通信のような応用に対する符号化ツールとそれに応する機能のさまざまな組み合わせを支援する。

【0004】 MPEG-4はマルチメディア環境でのビデオデータの効率的保存、送信及び操作を可能とする標準化された中核となる技術を与える。MPEG-4は効果的な圧縮、オブジェクト・スケーラビリティ、空間的及び時間的スケーラビリティ、及びエラー弾力性を達成する。

【0005】 MPEG-4ビデオ VM コーダ/デコーダ(code c)は移動補償を伴うブロックおよびオブジェクトベースのハイブリッドコーダである。重なりブロック移動補償を利用して、テクスチャが 8×8 離散コサイン変換(DCT)により符号化される。オブジェクト形状がアルファマップとして表され、内容ベース算術符号化(CAE)アルゴリズムまたは修正されたDCTコーダ(ともに時間的予測を使用する)を使って符号化される。コーダはコンピュータグラフィックで知られるスプライト(sprites)を扱う。ウェーブレット及びスプライトコーディングのような他の符号化方法がまた特別の応用に対して使用され得る。

【0006】 移動補償されたテクスチャのコード化がビデオコード化に関して周知の方法であり、3段階のプロセスとしてモデル化することができる。第1段階は、移動推定および補償(ME/MC)ならびに二次元(2-D)空間変換を含む信号処理である。ME/MC及び空間変換の目的は、複雑性の制約のもとでのエントロピーコード化および量子化の速度歪み性能を最適化するべくビデオシーケンス内の時間的及び空間的相関関係を利用することである。ME/MCに対する最も一般的な技術はブロックマッチングであり、最も一般的な空間変換はDCTであった。

【0007】

【発明が解決しようとする課題】 しかし、B-VOP内のマクロブロック(MB)がそれ自身インターレースされ、コード化され、かつ/またはインターレースされ、コード

化されている基準イメージを使用するとき、MBのME/MCに対して特別の関心が生じる。

【0008】特に、B-VOP内のMBに対し、移動ベクトル(MV)予測を与える効果的な技術が望まれている。またB-VOP内のフィールドコード化MBを直接的にモードコーディングする効果的な技術が望まれている。さらに、最も効果的なコーディングをもたらす基準イメージを選択すべく、フィールドコード化B-VOP内のMBに対して、コーディングモード決定処理が行われることが望まれている。本発明は上記及び他の利点を有する装置を与える。

【0009】

【課題を解決するための手段】本発明に従って、双方向に予測したビデオオブジェクト平面(VOP)における、現在のイメージ(たとえば、マイクロブロック)のようなデジタルビデオイメージのコード化する(特に、ここでは現在のイメージおよび/または現在のイメージをコード化するために使用した基準イメージがインターレース(たとえば、フィールド・コード化される)ための方法及び装置が提供される。

【0010】発明の第一の態様として、一連のデジタルビデオイメージにおいて、トップおよびボトムのフィールドをもつマクロブロック(MB)のような、現在の、双方向に予測した、フィールドコード化イメージに対して、直接モードの移動ベクトル(MV)を提供する方法である。トップおよびボトムのフィールドをもつ過去のフィールドコード化した基準イメージ、およびトップおよびボトムのフィールドをもつ未来のフィールドコード化した基準が決定される。未来のイメージは、 MV_{top} 、すなわち未来のイメージのトップフィールドのフォワードMVが前記過去のイメージのトップフィールドかボトムフィールドかのいずれかを基準とするように、過去のイメージを使用して予測される。基準としたフィールドは、未来のイメージのトップフィールドにおいて、MBに対し最適の適合MBを含む。

【0011】このMVは、それが過去のイメージ(たとえば、やがてはバックワード)を基準とするが、その予測が、過去のイメージから未来のイメージ(たとえば、やがてはフォワード)になるので、“forward(フォワード)”MVという。記憶方法として、予測の方向は、対応するMVの方向と反対となると思えばよい。

【0012】同様に、 MV_{bot} 、すなわち未来のイメージのボトムフィールドのフォワード移動ベクトルが過去のイメージのトップフィールドかボトムフィールドかのいずれかを基準とする。フォワードおよびバックワードMVは、未来のイメージの、対応するフィールドのフォワードMVをスケーリングすることにより、現在のイメージのトップおよび/またはボトムフィールドを予測するために決定される。

【0013】特に、 $MV_{f,top}$ 、すなわち現在のイメージのトップフィールドを予測するためのフォワード移動ベ

クトルは式、 $MV_{f,top} = (MV_{top} * TR_{B,top}) / TR_{D,top} + MV_D$ にしたがって決定され、ここで、 MV_D はサーチ領域に対するデルタ移動ベクトルであり、 $TR_{B,top}$ は現在のイメージのトップフィールドと、 MV_{top} により基準とされる過去のイメージのフィールドとの間の時間的な間隔に対応し、 $TR_{D,top}$ は未来のイメージのトップフィールドと、 MV_{top} により基準とされる過去のイメージのフィールドとの間の時間的な間隔に対応する。時間的な間隔は、イメージが表示されるフレームレートに関連付けられ得る。

【0014】同様に、 $MV_{f,bot}$ 、すなわち現在のイメージのボトムフィールドを予測するためのフォワード移動ベクトルは、式、 $MV_{f,bot} = (MV_{bot} * TR_{B,top}) / TR_{D,bot} + MV_D$ にしたがって決定され、ここで、 MV_D はサーチ領域に対するデルタ移動ベクトルであり、 $TR_{B,bot}$ は現在のイメージのボトムフィールドと、 MV_{bot} により基準とされる過去のイメージのフィールドとの間の時間的な間隔に対応し、 $TR_{D,bot}$ は未来のイメージのボトムフィールドと、 MV_{bot} により基準とされる過去のイメージのフィールドとの間の時間的な間隔に対応する。

【0015】 $MV_{B,top}$ 、すなわち現在のMBのトップフィールドを予測するためのバックワード移動ベクトルは式、 $MV_{B,top} = ((TR_{B,top} - TR_{D,top}) * MV_{top}) / TR_{D,top}$ (デルタ移動ベクトル、 $MV_D=0$ のとき)、または $MV_{B,top} = MV_{f,top} - MV_{top}$ ($MV_D \neq 0$ のとき)にしたがって決定される。

【0016】 $MV_{B,bot}$ 、すなわち現在のMBのボトムフィールドを予測するためのバックワード移動ベクトルは式、 $MV_{B,bot} = ((TR_{B,bot} - TR_{D,bot}) * MV_{bot}) / TR_{D,bot}$ (デルタ移動ベクトル、 $MV_D=0$ のとき)、または $MV_{B,bot} = MV_{f,bot} - MV_{bot}$ ($MV_D \neq 0$ のとき)にしたがって、決定される。対応するデコーダも与えられる。

【0017】発明の他の態様として、一連のデジタルビデオイメージにおいて、トップおよびボトムのフィールドをもつ現在の、予測したフィールドコード化MBに対して、コード化モードを選択する方法である。コード化モードは、バックワードモード(ここでは、基準MBは時間的に、表示順序で現在のMBの後となる)、フォワードモード(ここでは、基準MBは順序のMBの前となる)、またまたは平均(たとえば、双方向)(ここでは、先の基準MBおよび続く基準MBの平均が使用される)である。

【0018】本方法は、過去の基準MB(フォワードコード化モードに対応する)に關し、現在のMBに対する絶対差分エラーのフォワード合計、 $SAD_{forward,field}$ を決定する工程を含む。 $SAD_{forward,field}$ は、過去の基準MBにおける最適な適合MBと現在のMBとの間の画素ルミナンス値におけるエラーを示す。未来の基準MB(バックワードコード化モードに対応する)に關し、現在のMBに対する絶対差分エラーのフォワード合計、 $SAD_{backward,field}$ もまた決定される。 $SAD_{backward,field}$ は、未来の基準MB

における最適な適合MBと現在のMBとの間の画素ルミナンス値におけるエラーを示す。

【0019】未来および過去の基準MB（平均コード化モードに対応する）に関し、現在のMBに対する絶対差分エラーの平均合計、SAD_{average, field}もまた決定される。

SAD_{average, field}は、過去および未来の基準MBの最適な適合MBの平均のMBと現在のMBとの間の画素ルミナンス値におけるエラーを示す。

【0020】コード化モードはSADの最小にしたがって選択される。それぞれのコード化モードの必要なMVの数を説明するバイアス期間は、コード化モードの選択プロセスに分解され得る。

【0021】SAD_{forward, field}、SAD_{backward, field}、およびSAD_{average, field}は、トップおよびボトムのフィールドにわたってその成分期間を合計することにより決定される。

【0022】

【発明の実施の形態】本発明は双方向に予測されたビデオオブジェクトプレーン（B-BOP）においてマクロブロック（MB）のようなデジタルビデオイメージ（特に、MBおよび／またはMBをコード化するために使用される規準イメージがインターレース・コード化される場合）をコード化するための方法および装置を提供する。この方法はフィールドコード化した現在のMB（フレームコード化MB、ならびに必要な前方および公報PMVを含む）の頂部および底部フィールドに対して、予想移動ベクトル（PMV）を選択する方法を提供する。最適なモードを選択するために絶対的に差分期間の和に加え、フィールドコード化MBに対する直接コード化モードも与えられる。

【0023】図1は本発明に従うビデオ・オブジェクト・プレーン（VOP）コード化および復号化のプロセスを図示する。フレーム105が三つの画像エレメント（正方形前景エレメント107、長だ円前景エレメント108、および山形の景観エレメント109を含む）を含む。フレーム15において、エレメントは、VOP117が正方形前景エレメント107を表し、VOP118が長だ円エレメント108を表し、VOP119が山形の景観エレメント109を表すようにセグメンテーションマスクを使用して、指定されたVOPである。VOPは任意の形状を有し、VOPの連続がビデオオブジェクトとして知られる。完全な矩形のビデオフレームもまたVOPであると考えられる。したがって、“VOP”的用語はここでは任意の及び非任意（例えば矩形）のイメージ領域形状を示すのに使用される。セグメンテーションマスクが既知の技術を使用して得られ、ITU-R601輝度データのものに似たフォーマットを有する。各画素がビデオフレーム内のある領域に属するように識別される。

【0024】フレーム105およびフレーム115からのVOPデータは個別の符号化機能に供給される。特に、VOP117、118および119が、エンコーダ137、138および139においてそれぞれ、形状、移動およびテクスチャ符号化

を受ける。形状コード化とともに、バイナリーおよびグレースケール形状情報が符号化される。移動コード化とともに、形状情報はフレーム内の移動推定を使用してコード化される。テクスチャコード化とともに、DCTのような空間変換が、圧縮のためにコード化された可変長であり得る変換係数を得るために、実行される。

【0025】コード化されたVOPデータは次に、チャネル145にわたる伝送のために、マルチブレクサ（MUX）140で結合される。これにかわって、データは記録媒体上で記憶されてもよい。受信したコード化VOPデータはデマルチブレクサ（DEMUX）により分離され、分離したVOP117-118は復号化され、回復される。フレーム155、165および175は、VOP117、118および119がそれぞれ、復号化され、回復され、したがって、たとえばビデオライブラリー170とインターフェイスで接続するコンポジッタ160を使用して個別に操作され得ることを示す。

【0026】コンポジッタは、ユーザーの自宅に配置され、ユーザーが注文したイメージを得るために受信データを編集することができるパーソナルコンピューターのような装置であり得る。たとえば、ユーザーの個人的なライブラリー170は、受信したVOPと異なる、直前に保存したVOP178（例えば、円形）を含むことができる。ユーザーは、円形のVOP178を正方形のVOP117に置き換えたフレーム185を組み立てることができる。したがって、フレーム185は、受信したVOP118、119と、局部的に保存したVOP178とを含む。

【0027】他の例としては、背景のVOP109を、ユーザーの選択した背景に置き換えることができる。たとえば、テレビのニュース放送を見るときに、アナウンサーを、ニューススタジオのような背景から分離したVOPのようにコード化してもよい。ユーザーは、ライブラリー170から、又は株価や天気情報のチャンネルのような他のテレビ番組から背景を選択し得る。したがって、ユーザーは、ビデオ編集者であるかのように動作できる。

【0028】ビデオライブラリー170は、チャンネル145を介して受信されるVOPも保存することができ、また、インターネットのようなネットワークを介してVOPや他のイメージ要素にアクセスすることができる。一般に、ビデオセッションは、単一のVOP又はVOPのシーケンスを含む。

【0029】図1のビデオオブジェクトのコード化及び復号化のプロセスは、パーソナルコンピューターゲーム、仮想環境、グラフィカル・ユーザー・インターフェース、ビデオ会議、インターネット・アプリケーション等を含む多くのエンターテイメント、ビジネス、及び教育に応用できる。特に、本発明に従ったインターレースされ、コード化された（例えば、フィールドモード）VOPをもつME/MCの能力は、より大きな能力を与える。

【0030】図2は、本発明にしたがったエンコーダのプロック図である。エンコーダは、予測コード化VOP（P

-VOP) 及び双方向コード化VOP (B-VOP) の両方との使用に適している。

【0031】P-VOPは、イントラフレームモード又はインターフレーム・モードを使用して個々にコード化され得る多数のマクロブロック (MB) を含み得る。イントラフレーム (INTRA) のコード化とともに、マクロブロック (MB) は、他のMBを基準とせずに、コード化される。インターフレーム (INTER) コード化とともに、MBは、フォワード予測として知られるモードで時間的に継続するフレームに関して差分的にコード化される。アンカーフレーム (たとえば、VOP) は、P-VOP (B-VOPではない) でなければならない。I-VOPが予想してコード化されない内含した (たとえば、イントラコード化した) ブロックを含む。

【0032】フォワード予測とともに、現在のMBは、アンカーフレームのMBのサーチ領域と比較され、最良の適合を決定する。対応する移動ベクトル (MV) が、最良の適合のMBに対する現在のMBの相対変位を表す。さらに、P-VOPのためのアドバンスト予測モードが使用され、ここで、移動補償は、 16×16 のMBではなく、 8×8 のブロックで行われる。また、イントラフレームおよびインターフレームのコード化したP-VOPのMBの両方が、フレームモード又はフィールドモードでコード化され得る。

【0033】B-VOPは、バックワード予測、双方向予測、および直接モード (これら全てインターフレーム技術である) ならびに、P-VOPに関連して上述したようなフォワード予測モードを使用できる。B-VOPは、MPEG-4 MV8.0 の下で、イントラフレームコード化したMBを今は使用しないが、これは変化にしたがう。アンカーフレーム (たとえば、VOP) は、P-VOP (B-VOPではない) でなければならない。

【0034】B-VOPのバックワード予測とともに、現在のMBは、時間的に前にあるアンカーフレームのMBのサーチ領域と比較され、最良の適合を決定する。対応する移動MVが (フォワードMVとして知られている)、最良の適合のMBに対する現在のMBの相対変位を表す。B-VOPの双方向予測とともに、現在のMBは、時間的に前にあるアンカーフレーム及び時間的に続くアンカーフレームの両方のMBのサーチ領域と比較され、最良の適合を決定する。フォワード及びバックワード移動MVが、最良の適合のMBに対する現在のMBの相対変位を表す。さらに、平均的なイメージは、現在のMBをコード化するときに使用する最適の適合のMBから得られる。

【0035】後に続くP-VOPに配列されたマクロブロックが 8×8 のアドバンスト予測モードを使用すると、B-VOPの直接モード予測とともに、 8×8 のブロックのための移動ベクトルが導かれる。最良の適合のブロックを見つけ出すためのサーチを必要とせずに、B-VOPのブロックのための移動ベクトルを導くために、P-VOPの 8×8 のブロックの移動ベクトルが線形的にスケール (拡大または

縮小) される。

【0036】符号200で一般的に示されているエンコーダは、形状コーダ210と、移動推定機能220と、移動補償機能230と、テクスチャーコーダ240とを含み、各々は、ビデオ画素データ入力をターミナル205で受信する。移動推定機能220、移動補償機能230、テクスチャーコーダ240及び形状コーダ210はまた、MPEG-4のパラメータ "VOP_of_arbitrary_shape" のようなVOP形状情報入力もターミナル207で受信する。このパラメータがゼロであるとき、VOPは長方形の形状を有し、したがって、形状コーダ210は使用されない。

【0037】再構成されたアンカーバックワード機能250が、移動推定機能220及び移動補償機能230によって使用するための再構成されたアンカーバックワードを与える。現在のVOPが、テクスチャーコーダ240でエンコード化された留数 (residue) を与えるために、サブトラクタ260で移動補償されたアンカーバックワードから差し引かれる。テクスチャーコーダ240は、DCTにより、マルチプレクサ (MUX) 280にテクスチャ情報 (例えば、変換係数) を与えさせる。テクスチャーコーダ240はまた、再構成された前のアンカーバックワード機能250へ入力するための、加算器270で移動補償器230からの出力に加算される情報も与える。

【0038】移動情報 (例えば、移動ベクトル) は、移動推定機能220からMUX280へと与えられ、VOPの形状を示す形状情報は、形状コード化機能210からMUX280へ与えられる。MUX280は、対応する多重化したデータストリームを、データチャンネルにわたる継続した通信のためのバッファ290へ与える。

【0039】エンコーダへ入力される画素データは、YUV4:2:0フォーマットを有する。VOPは、境界長方形の手段によって表される。境界長方形の左上の座標は、最小の長方形の左上の座標よりも大きくなり最も近い偶数に丸められる。このことから、クロミナンス成分の境界長方形の左上の座標は、ルミナンス成分のものの半分である。

【0040】図3は、半分画素のサーチのための補間法を示す。移動推定及び移動補償 (ME/MC) は、一般に、現在のビデオフレームのブロック (例えば、現在のブロック) と基準フレームのサーチ領域にあるブロック (例えば、予測したブロック又は基準ブロック) とを適合させることを含む。予測コード化イメージに対して、基準ブロックは前のフレーム内にある。双方向予測 (B) コード化イメージに対して、前および続くフレーム内の予測ブロックが使用され得る。現在のブロックに対する基準ブロックの変位は、移動ベクトル (MV) であり、これは、水平 (x) 及び垂直 (y) 成分を有する。MV成分の正の値は、予測したブロックが現在のブロックの右方向及び下方向にあることを示す。

【0041】移動補償した差分ブロックが、予測したブロックの画素値を、現在のブロックのものから一点一点

13

差し引くことによって形成される。次に、テクスチャーコード化が、この差分ブロックで行われる。差分ブロックのコード化したMV及びコード化したテクスチャー情報は、デコーダへ伝送される。次に、デコーダは、MVに従って量子化した差分ブロックを、予測したブロックへ加えることによって、近似した現在のブロックを再構成することができる。ME/MCのブロックは、 16×16 のフレームブロック（マクロブロック）又は 8×8 のフレームブロック又は 16×8 のフィールドブロックであり得る。

【0042】MVの精度は、 $1/2$ 画素に設定される。補間法が、アンカーフレームで使用されなければならず、 $P(i+x, j+y)$ は、 x 及び y に対して定義され、整数の半分である。補間は、図3に示すように行われる。整数の画素の位置は、A, B, C及びDで示すよう符号“+”で表される。 $1/2$ 画素の位置は、a, b, c及びdで示すような円で示す。見てのとおり、 $a=A$, $b=(A+B)/2$, $c=(A+C)/2$, 及び $d=(A+B+C+D)/4$ であり、ここで、”//”は丸められる割り算を示す。補間の詳細は、前記のMPEG-4 VM8.0、ならびに“インターレースされたデジタルビデオ用のビデオオブジェクト平面の移動推定及び補償”と題する、この出願とともに譲渡された、エイフリングによる米国出願第08/897,847（1997年7月21日出願）（この文献はここに組み込まれる）に説明されている。

【0043】図6は、本発明に従った、適応フレーム／フィールド予測手法での画素ラインの再度の並び替えを示す。アドバンスト予測技術の第1の態様では、適応技術が、 16×16 の画素の現在のマクロブロック（MB）が、ME/MCコード化されるべきか、 8×8 の画素の四つのブロックの各々に分割すべきかを決定するために使用される。ここで、MBの画素ラインは、二つの 16×8 のフィールドブロックの同一のフィールドラインにグループ分けするために再度並び替えられ、 16×8 のブロックの各々 *

$$\text{If } \sum_{i=1}^4 SAD_8(MV_{xi}, MV_{yi}) < SAD_{16}(MV_x, MV_y) - 129,$$

ここで、NbはMBの不透明画素の数を示す。

【0048】インターレースしたビデオに対して、 $SAD_{top}(MV_{x_top}, MV_{y_top})$ 、及び $SAD_{bottom}(MV_{x_bottom}, MV_{y_bottom})$ を得る。ここで、 (MV_{x_top}, MV_{y_top}) 及び $(MV_{x_bottom}, MV_{y_bottom})$ は、トップ（偶数）及びボトム（bottom）フィールドのための移動ベクトルである。次に、フィールドの $1/2$ サンプルサーチから最小のSAD（例えば、 SAD_{top} 及び SAD_{bottom} のための）を有する基準フィールドを選択する。

$$(a) SAD_{16}(MV_x, MV_y), \quad (b) \sum_{i=1}^4 SAD_8(MV_{xi}, MV_{yi}) + 129,$$

$$\text{and } (c) SAD_{top}(MV_{x_top}, MV_{y_top}) + SAD_{bottom}(MV_{x_bottom}, MV_{y_bottom}) + 65.$$

【0050】 8×8 の予測が選択された場合、四つの 8×8 のルミナンスブロックに四つのMVがある（すなわち、8

14

*は、別々にME/MCコード化される。

【0044】フィールドの 16×16 のマクロブロック（MB）が、符号600で示されている。MBは、偶数番目のライン602, 604, 606, 608, 610, 612, 614, 616と、奇数番目のライン603, 605, 607, 609, 611, 613, 615, 617とを含む。よって、偶数及び奇数番目のラインは、インターリープされ、トップ及びボトム（又は、第1及び第2）のフィールドをそれぞれ形成する。

【0045】イメージ600の画素ラインが同一フィールドのルミナンスブロックを形成するために計算されると、符号650で示されるマクロブロックが形成される。符号645で示される矢印は、ライン602-617の再度の並び替えを示す。例えば、MB600の第1番目のラインである偶数番目のライン602は、MB650の第1番目のラインでもある。偶数番目のライン604は、MB650の第2番目のラインとなるように再度並び替えられる。同様に、偶数番目のライン606, 608, 610, 612, 614, 616は、それぞれ、MB650の第3番目から第8番目のラインとなるように再度並び替えられる。よって、偶数番目のラインをもつ 16×8 のルミナンス領域680が形成される。同様に、奇数番目のライン603, 605, 607, 609, 611, 613, 615, 617は、 16×8 の領域685を形成する。

【0046】P-VOPのためのMCモードを選択するための決定プロセスは、以下のとおりである。フレームビデオモードに対し、最初に、単一の 16×16 のブロック（例えば、 $SAD_{16}(MV_x, MV_y)$ ）及び四つの 8×8 のブロック（例えば、 $SAD_8(MV_{x1}, MV_{y1})$, $SAD_8(MV_{x2}, MV_{y2})$, $SAD_8(MV_{x3}, MV_{y3})$, 及び $SAD_8(MV_{x4}, MV_{y4})$ ）のための絶対差分和（SAD）を得る。下記の数1の場合、 8×8 の予測を選択し、それ以外は、 16×16 の予測を選択する。

【0047】

【数1】

※ 【0049】予測モード決定の全部は、以下の式2の最小のものを選択することに基づいている。式2の(a)が最小である場合、 16×16 の予測が使用される。式2の(b)が最小である場合、 8×8 の移動補償（アドバンスト予測モード）が使用される。数2の(c)が最小である場合、フィールドベースの移動推定が使用される。定数“65”はNb/4+1から得られる。

【数2】

*

15

$\times 8$ のブロックの各々にMVが一つづつある）。次に、二つのクロミナンスブロックのMVは、これら四つのMVの平均をとり、この平均値を二分することによって得られる。 8×8 のルミナンスブロックのMVの各々が $1/2$ 画素の精度を有することから、クロミナンスブロックのMVは $1/16$ 画素値を有し得る。下記の表1は、クロミナンスMVのための $1/2$ 画素値への $1/16$ 画素値の変換を示す。例えば、0から $2/16$ が0に丸められ、 $3/16$ から $13/16$ が $1/2$ に丸められ、 $14/16$ 及び $15/16$ が $2/2=1$ に丸められる。

表1

$1/16$ 画素値	0	2	3	4	5	6	7	8	9	10	11	12	13	14	15
$1/2$ 画素値	0	0	0	1	1	1	1	1	1	1	1	1	1	2	2

【0051】フィールド予測では、二つの 16×8 のブロックに二つのMVがある。ルミナンス予測は、以下のようにして生成される。マクロブロックの偶数番目のライン（例えば、ライン602, 604, 606, 608, 610, 612, 614, 616）は、特定された基準フィールドを使用するトップのフィールド移動ベクトルによって定義される。移動ベクトルは、全画素の垂直方向のずれが、垂直方向の移動ベクトルの座標の偶数の整数値に一致し、 $1/2$ 画素の垂直方向の変位が、奇数の整数値によって示されるように、フレームの座標に特定される。 $1/2$ 画素の垂直方向のオフセットが特定されると、同一の基準フィールド内のラインからの画素だけが組み合わされる。

【0052】二つのクロミナンスブロックのためのMVは、各々の成分を二分し、次に以下のようにして丸めることによって、（ルミナンス）移動ベクトルから導かれる。水平成分は、全ての分数値を $1/2$ 画素のオフセットにマッピングすることによって丸められる。垂直MBベクトル成分は、整数であり、得られるクロミナンスMVの垂直成分が整数に丸められる。二分して整数値が得られなかった場合、隣接の奇数に丸められる。ここで、奇数の整数値は、同一のフィールドのラインの間の垂直方向の補間を表す。

【0053】アドバンスト予測技術の第2の態様は、ルミナンスブロックのためのMCを重ね合わせが、MPEG-4 VM8.0およびエイフリング等に前記の文献に詳説されている。

【0054】B-VOPに対する特定のコード化技術がここで説明される。B-VOPのようなINTERコード化してVOPに対して、四つの予測モード、すなわち直接モード、補間（たとえば平均化または双方向）モード、バックワードモード、およびフォワードモードがある。後者の三つは非直接モードである。フォワードのみ、またはバックワードのみの予測がまた、“单方向”予測として知られている。B-VOPの予測したブロックは各モードに対してもいろいろに決定される。さらに、B-VOPのブロックおよびアンカーブロックがプログレッシブ（たとえばフレーム）・コード化され、またはインターレース（たとえば、フィールド）・コード化され得る。

【0055】一つのB-VOPが、異なるモードで予測された異なるMBをもつことができる。“B-VOP”という用語は、双方向に予測したブロックが含まれ得ることのみを示し、このことは要求されない。対照的に、P-VOPおよ

16

* 16 画素値を有し得る。下記の表1は、クロミナンスMVのための $1/2$ 画素値への $1/16$ 画素値の変換を示す。例えば、0から $2/16$ が0に丸められ、 $3/16$ から $13/16$ が $1/2$ に丸められ、 $14/16$ 及び $15/16$ が $2/2=1$ に丸められる。

【表1】

$1/16$ 画素値	0	2	3	4	5	6	7	8	9	10	11	12	13	14	15
$1/2$ 画素値	0	0	0	1	1	1	1	1	1	1	1	1	1	2	2

びI-VOPとともに、双方向に予測したMBは使用されない。

【0056】非直接モードB-VOP MBに対して、MVは差分的にコード化される。フォワードおよび双方向モードのフォワードMVに対し、ならびにバックワードおよび双方モードのバックワードMVに対し、同じ列の現在のMBの直前のMB “同じタイプ” のMV（たとえば、フォワードまたはバックワード）はプレディクターとして使用される。これは、ラスター順序の、一般的な伝送順序の直前MBと同じである。しかし、ラスター順序が伝送順序と異なるとき、伝送順序の直前のMVは、デコーダで、MBおよび対応するMVを記憶し、並べ替えする必要性を避けるために使用されるべきである。

【0057】同じタイプのMVを使用し、伝送順序がラスター順序と同じであり、ラスター順序が左から右に、上から下であると仮定して、左の隣接するMVのフォワードMVはB-VOPの現在のMBのフォワードMVに対し、プレディクターとして使用される。同様に、左の隣接するMBのバックワードMVはB-VOPの現在のMBのバックワードMVに対し、プレディクターとして使用される。現在のMBのMVは次に、プレディクターを使用して、差分的に符号化される。すなわち、現在のMBに対し決定されるMVとプレディクターとの間の差分は、移動ベクトルとして、デコーダに伝送される。デコーダでは、現在のMBのMVはPMVおよび差分MVを回復し、加えることにより決定される。

【0058】現在のMBがVOPの左端に位置する場合、現在のMBに対するプレディクターはゼロに設定される。

【0059】インターレース・コード化したB-VOPにおいて、トップおよびボトムのフィールドのそれぞれは、四つのMVの全体に対して二つの関連し予測移動ベクトルを有する。四つの予測MVは、送信順序で、前のアンカーモドのトップフィールドフォワードおよびボトムフィールドフォワード、ならびに次のアンカーモドのトップフィールドフォワードおよびボトムフィールドフォワードを表す。現在のMBおよびフォワードMB、および/または現在のMBおよびバックワードMBは、現在のMBのME/MCコード化に対して使用されない。B-VOPはINTRAコード化したMBを含まず、B-VOPの各MBはME/MCコード化されることになる。フォワードおよびバックワードアンカーモドはP-VOPまたはI-VOPからあり得て、フレームまたはフィールド・コード化され得る。

【0060】インターレースした、非直接モードのB-VOP

17

Pに対し、四つの可能な予測移動ベクトル(PMV)が以下の表2に示されている。表2の最初の欄は予測機能を示し、第二の欄はPMVのための指定子を示す。これらPMVは、いろいろなMB予測モードに対し、表3において示されているように使用される。

【0061】

【表2】

表2

予測機能	PMVタイプ
トップフィールド、フォワード	0
ボトムフィールド、フォワード	1
トップフィールド、バックワード	2
ボトムフィールド、バックワード	3

【表3】

マクロブロック	使用PMVタイプ
フレーム、フォワード	0
フレーム、バックワード	2
フレーム、双方向	0, 2
フィールド、フォワード	0, 1
フィールド、バックワード	2, 3
フィールド、双方向	0, 1, 2, 3

【0062】たとえば、表3はフォワードモード(たとえば、“フィールド、フォワード”)をもつ、現在のフィールドモードMBに対して、トップフィールドフォワード(“0”)およびボトムフィールドフォワード(“0”)移動ベクトルプレディクターが使用される。

【0063】差分コード化において使用された後、現在のMBの移動ベクトルは伝送順序で、続くMBに対しPMVとなる。PMVはMBの各列の始まりでゼロにリセットされる。その理由は先の列の終わりにおけるMBのMVは現在の列の始まりにおいてMBのMVと近似したものとなりそうもないからである。プレディクターはまた直接モードのMBに対して使用されない。飛び越したMBに対し、PMVは最後の値を保持する。

【0064】B-VOPの直接モードのコード化でもっては、ベクトル差分は伝送されない。代わって、フォワー*

表4 (予測移動ベクトルインデックスpmv[])

伝送順序の 先のマイク ロブロック タイプ	現在のマクロブロックタイプ							
	直接	フォワード	バック	平均	フォワード	バック	平均	モード
フレームモード	フレームモード	フレームモード	フレームモード	フレームモード	フレームモード	フレームモード	フレームモード	フレームモード
更新するためのpmv[0] non	0, 1	2, 3	0, 1, 2, 3	0, 1	2, 3	0, 1, 2, 3	0, 1, 2, 3	0, 1, 2, 3
使用するためのpmv[0] non	0	2	0, 2	0, 1	2, 3	0, 1, 2, 3	0, 1, 2, 3	0, 1, 2, 3

【0068】表4が、単に、現在のMBに対する予測MVを

*ドおよびバックワードMVは、時間的に次のP-VOP MBのMVから、一つのデルタMVにより関連をもって、デコーダにおいて直接計算される。このデルタMVはMVデータが伝送されないことからこの技術は効果的である。

【0065】表4は、PMVが先および現在のMBタイプに基づいた現在のB-VOPの移動ベクトルをコード化するために使用されることを要約して示す。B-VOPに対し、予測移動ベクトル、pmv[]の列が与えられ、ゼロから三つ(たとえば、pmv[0], pmv[1], pmv[2]およびpmv[3])

10 に割り付けられる。このインデックスpmv[]は伝送されるが、MVコード化タイプおよび復号かされる特別なベクトルにしたがって使用するために、pmv[]インデックスを決定できる。B-VOPをコード化した後に、PMVベクトルのいくつかは現在のMBの移動ベクトルと同じになるよう更新される。最初の一つ、二つまたは四つのPMVは現在のMBと関連したMVの数に依存して更新される。

【0066】たとえば、フォワード、フィールド予測したMVが二つの移動ベクトルを有する。ここで、pmv[0]はトップフィールド、フォワードに対しPMVであり、pmv

20 [0]はボトムフィールド、フォワードに対しPMVである。バックワード、フィールド予測したMBに対して、pmv[2]はトップフィールドバックワードに対しPMVである。双方向、フィールド予測したMBに対し、pmv[0]はトップフィールドフォワードに対してPMVであり、pmv[1]はボトムフィールドフォワードに対しPMVであり、pmv[2]はトップフィールドバックワードに対しPMVであり、pmv[3]はボトムフィールドバックワードに対しPMVである。

30 フォワードまたはバックワード予測したフレームモードB-VOPMBに対し、唯一のMVがあり、pmv[0]のみがフォワードに対し使用され、pmv[2]がバックワードに対して使用される。平均(たとえば、双方向)予測したフレームモードB-VOPMBに対し、二つのMV、すなわちフォワードMVに対するpmv[0]およびバックワードMVに対するpmv[2]がある。列指定した“更新するためのpmv[]”は、一つ、二つまたは四つのMVが更新されたかどうかを示す。

【0067】

【表4】

50 選択すべく、本発明の技術を実行するための速記表記で

19

あることは分かるであろう。しかし、このようなものは他のいろいろな方法で表すことができる。

【0069】INTRAブロックDC適応予測が、`dct_type`の値に拘わらず、MPEG-4MV8.0に説明されているのと同じアルゴリズムを使用することができる。INTRAブロック適応予測が、最初の列の係数が上記コード化されたブロックからコピーされるべきときを除いて、MPEG-4MV8.0に説明されているように実行される。この操作は、`dct_type`が現在のMBおよび上記ブロックに対して同じ値をもつときのみ可能となる。`dct_type`が異なると、AC予測が最初の欄をブロックから左側にコピーすることによってのみ起こる。左側ブロックがないとき、ゼロがAC予測に対して使用される。

【0070】図4は本発明にしたがってインターレース・コード化したB-VOPのトップフィールドの直接モードのコード化を図示する。現在のMBと同じ相対的位置（たとえば、共同位置づけ）にある未来のアンカー画像が、(1) 16×16(フレーム) MB、(2) イントラMBまたは(3) 8×8(先行予想)のようにコード化されるときには、現在のマイクロブロック(MB)に対して、プログレッシブ・直接コード化モードが使用される。

【0071】直接モード予想は、共同位置づけした未来のアンカー画像MBがインターレースしたMBとしてコード化されたときには、インターレースされる。直接モードは、そのバイアスされたSADが全てのB-VOP MV予想の最小であるときに、現在のMBをコード化するために使用される。インターレース・コード化したMBに対する直接モードが、現在のトップおよびボトムのフィールドに対して別々に、予測MBを形成する。双方向フィールド移動補償したMB（たとえば、トップフィールドフォワード、ボトムフィールドフォワード、トップフィールドバックワード、およびボトムフィールドバックワード）の四つのフィールド移動ベクトル(MV)は、未来のアンカー画像の対応するMBのそれぞれのMVから直接計算される。

【0072】その技術は、必要としたサーチが非常に減少し、伝送したMVデータの量が減少したので、十分である。MVおよび基準フィールドが決定されると、現在のMBは、双方向フィールド予想したMBであると考えられる。唯一のデータ移動ベクトル(両フィールドに対して使用されたもの)は、フィールド予測したMBに対するビットストリームにおいて生じる。

【0073】現在のMBのトップフィールドに対する予測は、未来のアンカー画像（これは、 $MV=0$ でI-VOPまたはP-VOPであり得る）のMBのトップフィールド、および未来のアンカーMBのトップフィールドの対応するMVにより選択される先行のアンカー画像の過去の基準フィールドに基づく。すなわち、現在のMBへと対応して位置づけられる未来のアンカー画像のトップフィールドMBは過去のアンカー画像のトップフィールドまたはボトムフィールドのいずれかで、最適な適合をもつ。この最適な適合MBは

20

次に、現在のMBのトップフィールドに対するアンカーMBとして使用される。包括的なサーチが、アルファ移動ベクトルMVが与えられる、MB基礎によりMBにおいて共同位置づけした未来のアンカーMVを決定するために使用される。

【0074】現在のMBのボトムフィールドに対する移動ベクトルが、同様に、（過去のアンカー画像のトップまたはボトムのフィールド内の最適な適合を順に基準とする）未来のアンカーMBの、対応して位置づけられたボトムフィールドのMVを使用して決定される。

【0075】基本的に、トップフィールドベクトルは、(a) 対応して位置づけした未来のアンカーMBのトップフィールドから得た画素と (b) 対応して位置づけしたトップフィールドMVにより基準とした過去のアンカーフィールドからの画素との平均であるMVプレディクターを構成するために使用される。同様に、ボトムフィールド移動ベクトルは、(a) 対応して位置づけした未来のアンカーMBのボトムフィールドから得た画素と (b) 対応して位置づけられたボトムフィールドMVにより基準とした過去のアンカーフィールドからの画素との平均であるMVプレディクターを構成するために使用される。

【0076】図4に示したように、現在のB-VOP MV420はトップフィールド430およびボトムフィールド425を含み、過去のアンカーVOP MB400はトップフィールド410およびボトムフィールド405を含み、未来のアンカーVOP MB440はトップフィールド450およびボトムフィールド445を含む。

【0077】 MV_{top} は、過去のアンカーMB400の最適な適合MBを示す未来のアンカーMB440のトップフィールド450に対するフォワード移動ベクトルである。 MV_{top} は先のイメージ（たとえば、時間的にバックワード）を基準としているとしても、未来のアンカーVOP440が過去のアンカーVOP400に関して時間的にフォワードであるので、それはフォワードMVである。この例においてトップフィールド410かボトムフィールド405かのいずれかが基準とされ得るが、 MV_{top} は過去のアンカーMB4000のボトムフィールド405を基準とする。 $MV_{f,top}$ は現在のMBのトップフィールドのフォワードMBであり、 $MV_{b,top}$ は現在のMBのトップフィールドのバックワードMVである。画素データが、デコーダにおいて、 $MV_{b,top}$ および $MV_{f,top}$ によりそれぞれ同定される未来および過去のアンカーアイメージの画素データを平均化することにより、そして伝送される留数で平均化されたイメージの合計をとることにより、導かれる。

【0078】トップフィールドに対する移動ベクトルは以下のように計算される：

$MV_D=0$ のとき、

$$MV_{f,top} = (TR_{b,top} * MV_{top}) / TR_{D,top} + MVD$$

$$MV_{b,top} = ((TR_{b,top} - TR_{D,top}) * MV_{top}) / TR_{D,I}$$

50 $MV_D \neq 0$ のとき、

21

$$MV_{b,top} = (MV_f,top - MV_{top})$$

MVD はデルタ、またはオフセット、移動ベクトルである。移動ベクトルが二次元あることに注意されたい。さらに、移動ベクトルは整数の1/2ルマ移動ベクトルである。スラッシュ "/" は整数の分数をゼロに向けてトランケートすることを示す。また、未来のアンカー-VOPは常に、フィールド直接モードに対してP-VOPである。未来のアンカーがI-VOPであったとき、MVはゼロとなり、 16×16 プログレッシブ直接モードは使用される。 $TR_{B,top}$ は、過去の基準フィールド（たとえば、トップまたはボトム）（この例においては、ボトムフィールド405である）と現在のB-VOP420のトップフィールド430との間のフィールドの時間的間隔である。 $TR_{D,top}$ は、過去の基準フィールド（たとえば、トップまたはボトム）（この例においては、ボトムフィールド405である）と未来のトップフィールド450との間の時間的間隔である。

【0079】図5は、本発明にしたがったインターレース・コード化したB-VOPのボトムフィールドの直接モードのコード化を図示する。ソースをインターレースしたビデオが最初にトップフィールドまたは最初にボトムフィールドのフォーマットを有し得ることに注目すべきである。ボトムフィールドの第一のフォーマットが図4および図5に示されている。図4と同様の符号が付されている。ここでは、移動ベクトル MV_{bot} は、過去のアンカーメンブロック(MB)400の最適な適合MBを示す未来のアンカーメンブロック(MB)440のボトムフィールド445に対するフォワード移動ベクトルである。この例において、トップフィールド410かボトムフィールド405かのいずれかが基準とされるが、 MV_{bot} は、過去のアンカーメンブロック(MB)405を基準とする。

【0080】ボトムフィールドに対する移動ベクトルは、次のように、トップフィールド移動ベクトルとパラレルな方法で計算される：

表5-時間的補正、 δ					
基準フィールド		最初にボトムフィールド		最初にトップフィールド	
未来の アンカー	未来の アンカー	トップフィールド	δ	ボトムフィールド	δ
トップ	トップ	0	-1	0	1
トップ	ボトム	0	0	0	0
ボトム	トップ	1	-1	-1	1
ボトム	ボトム	1	0	-1	0

【0085】効果的なコード化のために、適当なコード化モード決定プロセスが必要となる。示されているように、B-VOPに対して、MBは(1)直接モード、(2)補償された 16×16 移動(フォワード、バックワード、および平均モードを含む)、または(3)フィールド移動補償(フォワード、バックワード、および平均モードを含む)を使用してコード化され得る。現在のMBのフレーム

22

* $MVD=0$ のとき、

$$MV_{f,bot} = (MV_B,bot * MV_{bot}) / TR_{D,bot} + MVD$$

$$MV_{b,bot} = ((MR_B,bot - MR_D,bot) * MV_{bot}) / TR_{D,I}$$

$MVD \neq 0$ のとき、

$$MV_{b,bot} = (MV_{f,bot} - MV_{top})$$

$TR_{B,bot}$ は、過去の基準フィールド（たとえばトップまたはボトム）（この例においては、ボトムフィールド405）と、現在のB-VOP420のボトム425との間の時間的な間隔である。 $TR_{D,bot}$ は、過去の基準フィールド（たとえばトップまたはボトム）（この例においては、ボトムフィールド405）と、未来のボトム基準フィールド445との間の時間的な間隔である。

【0081】図4および5の例において、 $TR_{B,top}$ 、 $TR_{D,top}$ 、 $TR_{B,bot}$ 、 $TR_{D,bot}$ の計算は現在のフィールド、基準フィールド、およびフレーム時間基準のみならず、現在のビデオが最初にトップフィールドであるが、最初にボトムフィールドであるかどうかに依存する。

【0082】

$$TR_{D,top} \text{ または } TR_{D,bot} = 2 * (TR_{furre} - TR_{past}) + \delta,$$

$$TR_{B,top} \text{ または } TR_{B,bot} = 2 * (TR_{current} - TR_{past}) + \delta$$

TR_{furre} 、 $TR_{current}$ 、および TR_{past} は、表示の順序で、それぞれ未来、現在および過去のフレームのフレーム数であり、フィールド間の時間的な間隔への付加的な補正 δ は以下表5に示されている。 δ はフィールド間隔の単位である。

【0083】たとえば、第一の欄の差異母の列の指定 "1" は未来のアンカーフィールドがトップフィールドであり、基準となったフィールドがボトムフィールドであることを示す。このことは図4に示されている。指定 "1" は、未来のアンカーフィールドがボトムフィールドであり、基準となったフィールドもボトムフィールドであることを示す。このことは図5に示されている。

【0084】

【表5】

*

またはフィールド直接コード化が、対応する未来のアンカーメンブロック(MB)がそれぞれ、フレームまたはフィールド直接コード化されるときに使用される。

【0086】B-VOPのフィールド移動補償したMBに対して、復号化したアンカーフィールドに関する、最小のルミナンス1/2画素SADに基づくフォワード、バックワード、または平均モードでMBをコード化するために、決定がなされ

る。特に、七つのバイアスしたSADタームが次のように計算される。

- (1) SAD_{direct}+b1
- (2) SAD_{forward}+b2
- (3) SAD_{backward}+b3
- (4) SAD_{average}+b4
- (5) SAD_{forward, field}+b3
- (6) SAD_{backward, field}+b3
- (7) SAD_{average, field}+b4

ここで、添え字は、直接モード、フォワード移動予測、バックワード移動予測、平均（すなわち、補間または双方向）移動予測、フレームモード（すなわち、局所的にプログレッシブ）およびフィールドモード（すなわち、局所的にインターレース）を示す。上記フィールドSAD（すなわち、SAD_{forward, field}、SAD_{backward, field}、SAD_{average, field}）はトップおよびボトムフィールドSADの合計で、それぞれはそれ自身の基準フィールドおよび移動ベクトルもつ。特別に、SAD_{forward, field}=SAD_{forward, top field} + SAD_{forward, bottom field}; SAD_{backward, field}=SAD_{backward, top field}+SAD_{backward, bottom field}

* 20

表6

モード	移動ベクトルの数	bi	バイアス	値
直接	1	b1	- (Nb/2+1)	-129
フレームフォワード	1	b2	0	0
フレームバックワード	1	b2	0	0
フレーム平均	2	b3	(Nb/4+1)	65
フィールドフォワード	2	b3	(Nb/4+1)	65
フィールドバックワード	2	b3	(Nb/4+1)	65
フィールド平均	4	b4	(Nb/2+1)	129

直接モードに対する負のバイアスはプログレッシブビデオに対する現存のMPEG-4と整合し、相対的により飛び越えたMBをもたらす。

【0090】図7は、本発明に従ったデコーダのブロック図である。符号700で一般的に示したデコーダは、図2のエンコーダから伝送された符号化したデータ信号を受信し復号化するために使用される。符号化したビデオイメージデータ及び差分的に符号化した移動ベクトル(MV)データは、ターミナル740で受信され、デマルチプレクサ(DEMUX)742へ与えられる。符号化したビデオイメージデータは、典型的に、予測エラー信号(例えば、留数)のようなDCT変換係数に差分的に符号化される。

【0091】移動補償機能750及びVOP再構成機能752へ、順に与えられる形状情報を受信するために、VOPが任意の形状を有するとき、形状復号化機能744が、データを処理する。留数情報を回復するために、テクスチャ復号化機能746が、変換係数で逆DCTを行う。INTRAコード化したマクロブロック(MB)のため、画素情報が、直接に回復され、VOP再構成機能752へ与えられる。

【0092】B-VOPのこれらのような、INTERコード化

* backward, bottom field; SAD_{average, field}=SAD

average, top field+SAD_{average, bottom field}

【0087】SAD_{direct}は最適な直接モード予測であり、SAD_{forward}はフォワード(過去の)基準からの最適な16×16予測であり、SAD_{backward}はバックワード(未来の)基準からの最適な16×16予測であり、SAD_{average}は最適なフォワードおよび最適なバックワード基準の画素対画素平均により形成された最適な16×16予測であり、SAD_{forward, field}はフォワード(過去の)基準からの最適な16×16予測であり、SAD_{backward, field}はバックワード(未来の)基準からの最適な16×16予測であり、SAD_{average, field}は最適なフォワードおよび最適なバックワード基準の画素対画素平均により形成された最適な16×16予測である。

【0088】biは、より移動ベクトルを必要とする予測モードを説明するために、以下の表6により定義されるバイアス値である。

【0089】

【表6】

30 したブロック及びMBに対し、テクスチャ復号化機能746から再構成VOP機能752へ与えられた画素情報は、現在のマクロブロックと基準イメージとの間の留数を表す。基準イメージはフォワードまたはバックワードMVにより示される一つのアンカーメモリからの画素データであってもよい。これに代わって、補間した(たとえば、平均化した)MBに対し、基準イメージは二つの基準MB、たとえば、一つが過去のアンカーメモリ、一つが未来のアンカーメモリからの画素データの平均である。この場合、デコーダは、現在のMB画素データを受信する前に、フォワードおよびバックワードMVにしたがって平均化した画素データを計算しなければならない。

【0093】INTERコード化したブロック及びMBに対し、移動復号化機能748が、符号化したMVデータを処理し、差分移動MVを回復し、そしてそれらを移動補償機能750及びRAMのような移動ベクトルメモリー749へ与える。移動補償機能750は、本発明に従って、差分移動MVを受信し、基準移動ベクトル(たとえば、移動ベクトル予測値、またはPMV)を決定する。PMVは、コード化モード(たとえば、フォワード、バックワード、双方向または直接)にしたがって決定される。

【0094】一旦、移動補償機能750が全基準移動MVを決定し、それに現在のマクロブロックの差分移動ベクトルを加えると、現在のMBの全MVは利用可能である。このことから、移動補償機能750は、RAMのようなVOPメモリ-754からのアンカーフレームの最良の適合データを検索し、必要ならば平均化したイメージを計算し、アンカーフレームの画素データをVOP再構成機能へ与え、現在のMBを再構成する。

【0095】検索し、または計算した最良の適合データは、VOP再構成機能752で画素の留数へ反対に加えられ、復号化した現在のMB又はブロックを得る。再構成したブロックは、ビデオ入力信号として出力され、VOPメモリ-754にも与えられ、新しいアンカーフレームデータを与える。ここで、B-VOPに対するアンカーフレームが表示順序で時間的に未来のフレームまたはフィールドなので、適当なビデオデータバッファ能力が、フレーム伝送及び表示順序に従って要求される。

【0096】図8は、本発明に従ったMBパケット構造を示す。この構造は、B-VOPに適しており、デコーダにより受信されたデータのフォーマットを示す。パケットは便宜上四つの列で示されていることを理解されたい。パケットは実際にはトップの列から開始し、列の左から右へと、連続して伝送される。第1のパケット810は、first_shape_code, MVD_shape, CR, STおよびBACといったフィールドを含む。第2のパケット830は、MODBおよびMBTYPEといったフィールドを含む。第3のパケット850は、CBPB, DQUANT, Interlaced_information, MVD_f, MVD_bおよびMVDBといったフィールドを含む。第4のパケットは、CODE, CBPBA, Alpha Block DataおよびBlock Dataといったフィールドを含む。上記のフィールドの各々は、MPEG-4標準規格に従って定義される。

【0097】first_shape_codeはMBがVOPの境界ボックスにあるかどうかを示す。CRはBinary Alpha Blockに対する変換比を示す。STは水平または垂直走査順序を示す。BACは二進算術コードワードを示す。

【0098】MODB (MBのモードを示す) は B-VOPのすべてのコード化した(飛び越しなし) MBに対してある。差分移動ベクトル (MVD_f, MVD_b, MVDB) およびCBPBはMDBにより示されるならばある。マクロブロックタイプがMBTYPE (移動ベクトルモード (MVD) および量子化 (DQUANT) を知らせる) により示される。インターレースモードでは、MB当たり四つのMVである。MBTYPEはコード化タイプ、たとえばフォワード、バックワード、双方向または直接を示す。CBPBはBタイプのマクロブロックに対するCoded Block Patternである。CBPBAは、同様に四ビットの最大をもつことを除き、DBPBと定義される。DQUANTは量子化器の値の変化を定義する。

【0099】第3のパッケージ850のフィールドInterlaced_informationは、MBがインターレース・コード化されるかどうかを示し、現在のMB又はブロックのコード化

モードのデコーダに情報を与えるフィールドMV基準データを与える。デコーダは、現在のMBのためのMVを計算するのにこの情報を使用する。Interlaced_informationフィールドは、デコーダの移動ベクトルメモリー749又は他のメモリーで要求されるように継続して使用するためには保存され得る。

【0100】Interlaced_informationフィールドは、また、図6に関連して上述したように、フィールドコード化したMBのトップ及びボトムのフィールド画素ラインが、インターリープした順序から再並び替えられるかどうかを示すflag dct typeを含む。

【0101】示したMBの層構造は、VOP_prediction_type==10のとき使用される。CODが最近の復号化したI-またはP-VOPのMBに対し飛び越し (COD=="1") 示すとき、B-VOPの共同配置づけした(たとえば、共同位置づけした) MBはまた飛び越している。すなわち、ビットストリームにはいかなる情報も含まれない。

【0102】MVD_fは時間的に先の基準VOP (I-またはP-VOP) に関するB-VOPのMBの移動ベクトルである。それは、垂直成分に対する可変長のコードワードが続く、垂直成分に対する可変長のコードワードから成る。“1”のfield predictionおよびフォワードのMBTPEまたはinterpolateをもつインターレスしたB-VOP MBに対して、MVD_fは過去のアンカーボックスを基準とする一対のフィールド移動ベクトル (ボトムフィールドが続くトップフィールド) を表す。

【0103】MVD_bは、時間的に続く基準VOP (I-またはP-VOP) に関するB-VOPのMBの移動ベクトルである。それは、垂直成分に対する可変長のコードワードが続く、垂直成分に対する可変長のコードワードから成る。“1”的field predictionおよびフォワードのMBTPEまたはinterpolateをもつインターレスしたB-VOP MBに対して、MVD_bは未来のアンカーボックスを基準とする一対のフィールドMV (ボトムフィールドが続くトップフィールド) を表す。

【0104】MVDBは単に、直接モードがMODBおよびMBYPEにより示されるときB-VOPにあり、各ベクトルの垂直成分に対する可変長のコードワードが続く水平成分に対する可変長のコードワードから成る。MVDBはP-VOP MB移動ベクトルをスケーリングすることにより得られるB-VOP MBを補正するために使用される。

【0105】CODAはグレースケール形状コード化として参照される。

【0106】図8に示した配列は単に一実施例であり、適切をデコーダに通信する他の種々の配列は当業者には明らかであろう。

【0107】本発明にしたがって使用するためのビットストリームのシンタックスおよびMB層シンタックスがエイフリング等の前記出願およびMPEG-4 VM8.0に説明されている。例えば、マクロブロックがフィールドコード化

されるかどうかといった、マクロブロックのコード化モードを決定する。

【0108】したがって、特に現在のMBがフィールドコード化され、および/またはアンカーMBがフィールドコード化されるとき、本発明がB-VOPの現在のMBを復号化するための方法を提供する。フィールドコード化したMBに対して直接コード化するための方法が、最適のモードを選択するために絶対的な差分タームの合計の最小を使用するコード化決定プロセスに加えて、表される。予測移動ベクトル(PMV)がまた、必要なフォワードおよびバックワードPMVを含むフィールドコード化した現在のMBのトップおよびボトムフィールドに、さらにフレームコード化したMBに与えられる。

【0109】本発明は、様々な特定的な実施例に関連して説明されたが、多種の適用及び変更が、特許請求の範囲に記載の本発明の思想及び範囲から逸脱せずになされることは、当業者には明らかである。

【図面の簡単な説明】

【図1】本発明に従ったビデオオブジェクト平面(VOP)のコード化及び復号化プロセスを示す。

【図2】本発明に従ったエンコーダのブロック図であ

*る。

【図3】1/2画素サーチのための補間法を示す。

【図4】本発明に従って、インターレース・コード化したB-VOPのトップフィールドの直接モードコード化を示す。

【図5】本発明に従って、インターレース・コード化したB-VOPのボトムフィールドの直接モードコード化を示す。

【図6】本発明に従って、適用フレーム/フィールド予測手法での画素ラインの再並び替えを示す。

【図7】本発明に従ったデコーダのブロック図を示す。

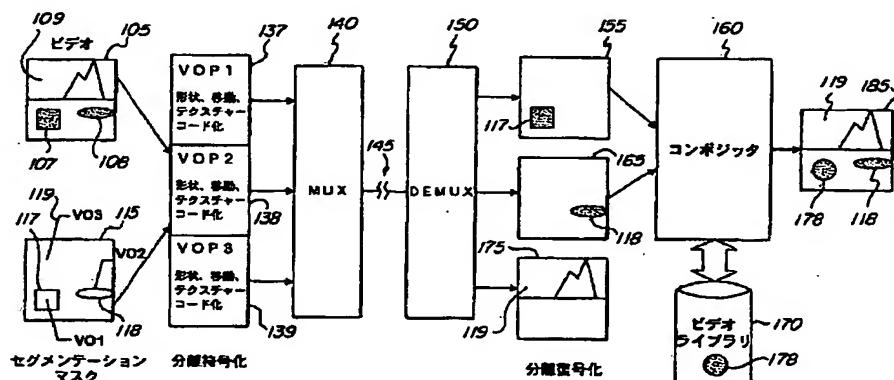
【図8】本発明に従ったマクロブロック層構造を示す。

【符号の説明】

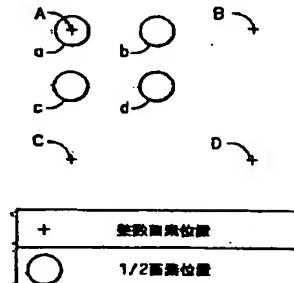
400	基準イメージ
410	基準フィールド
405	基準フィールド
420	ビデオオブジェクト平面
425	ボトムフィールド
430	トップフィールド
20 440	基準イメージ

20

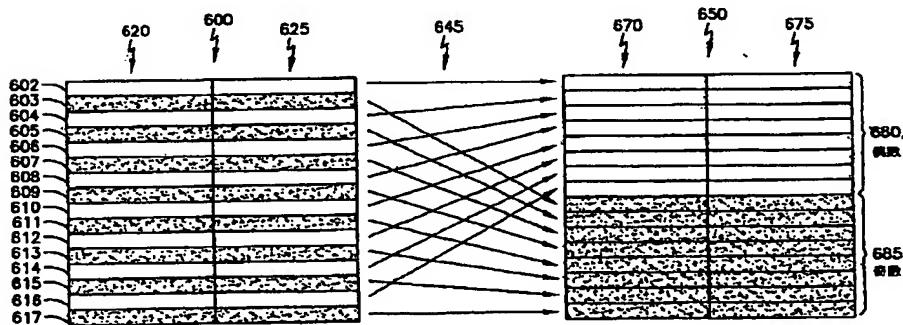
【図1】



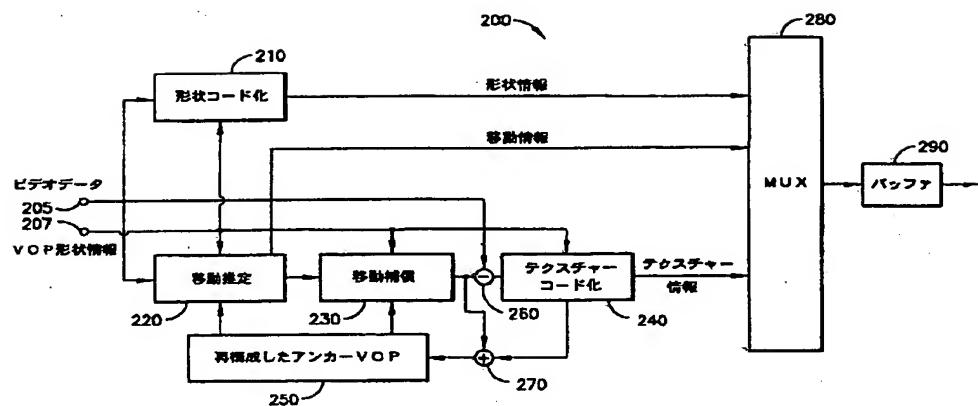
【図3】



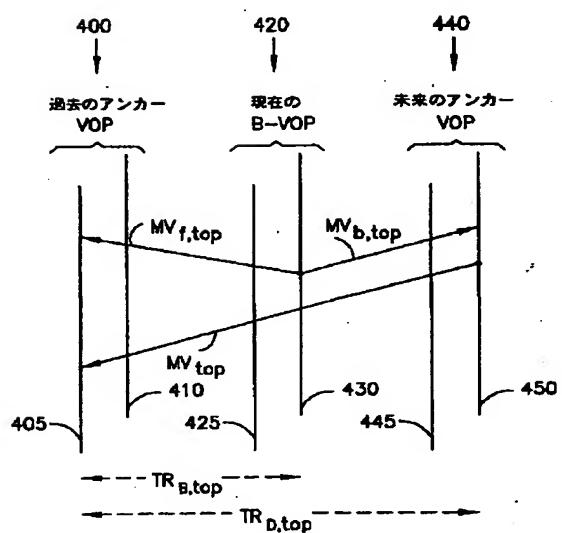
【図6】



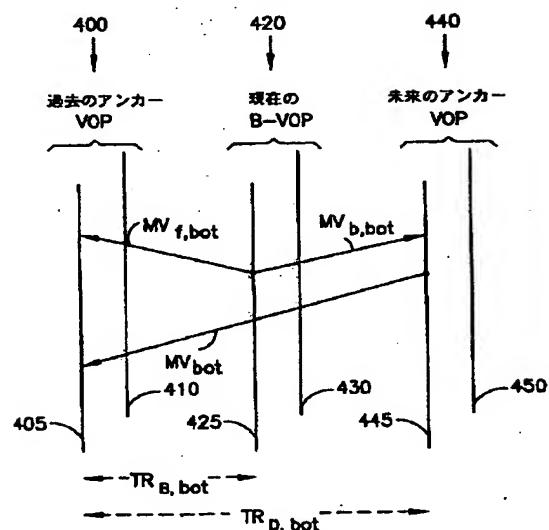
【図2】



【図4】



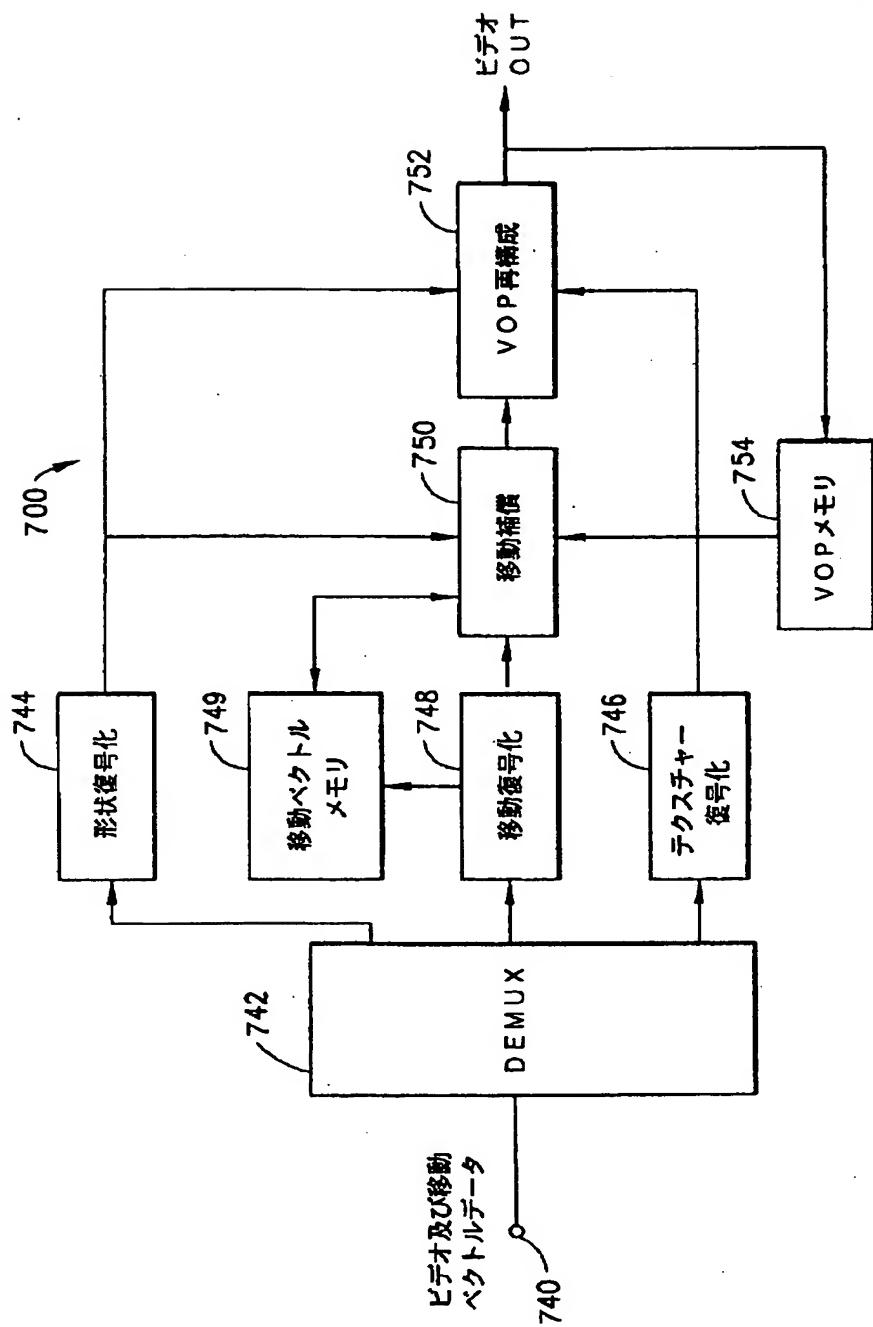
【図5】



【図8】

810 →	First_shape_code	MVD_shape	CR	ST	BAC
830 →	MODB	MBTYPE			
850 →	CBPB	DQUANT	Interlaced_information	MVD _f	MVD _b
870 →	CODA	CBPA	Alpha Block Data	Block Data	

【図7】



フロントページの続き

(72)発明者 スーミン・チェン
 アメリカ合衆国カリフォルニア州サンディ
 エゴ、フォックスクロフト プレイス
 8560

(72)発明者 アジャイ・ルスラ
 アメリカ合衆国カリフォルニア州サンディ
 エゴ、スチバ コート 8714

【外國語明細書】

1. Title of Invention

PREDICTION AND CODING OF BI-DIRECTIONALLY PREDICTED VIDEO
OBJECT PLANES FOR INTERLACED DIGITAL VIDEO

2. Claims

1. A method for calculating direct mode motion vectors for a current bi-directionally predicted, field coded image having top and bottom fields, in a sequence of digital video images, comprising the steps of:

determining a past field coded reference image having top and bottom fields, and a future field coded reference image having top and bottom fields;

wherein the future image is predicted using the past image such that MV_{top} , a forward motion vector of the top field of the future image, references one of the top and bottom fields of the past image, and MV_{bot} , a forward motion vector of the bottom field of the future image, references one of the top and bottom fields of said past image; and determining forward and backward motion vectors for predicting at least one of the top and bottom fields of the current image by scaling the forward motion vector of the corresponding field of the future image.

2. The method of claim 1, wherein:

MV_{top} , the forward motion vector for predicting the top field of the current image is determined according to the expression $(MV_{top} * TR_{B,top}) / TR_{D,top} + MV_D$;

where $TR_{B,top}$ corresponds to a temporal spacing between the top field of the current image and the field of the past image which is referenced by MV_{top} , $TR_{D,top}$ corresponds to a temporal spacing between the

top field of the future image and the field of the past image which is referenced by MV_{top} , and MV_b is a delta motion vector.

3. The method of claim 2, wherein:

$MV_{f,top}$ is determined using integer division with truncation toward zero; and

MV_{cep} and MV_{bot} are integer half-luma pel motion vectors.

4. The method of claim 2 or 3, wherein:

$TR_{s,top}$ and $TR_{b,top}$ incorporate a temporal correction which accounts for whether said current field coded image is top field first or bottom field first.

5. The method of one of the preceding claims, wherein:

$MV_{f,bot}$, the forward motion vector for predicting the bottom field of the current image is determined according to the expression $(MV_{bot} * TR_{s,bot}) / TR_{b,bot} + MV_b$;

where $TR_{b,bot}$ corresponds to a temporal spacing between the bottom field of the current image and the field of the past image which is referenced by MV_{bot} , $TR_{b,bot}$ corresponds to a temporal spacing between the bottom field of the future image and the field of the past image which is referenced by MV_{bot} , and MV_b is a delta motion vector.

6. The method of claim 5, wherein:

$MV_{f,bot}$ is determined using integer division with truncation toward zero; and

MV_{top} and MV_{bot} are integer half-luma pel motion vectors.

7. The method of claim 5 or 6, wherein:

$TR_{b,bot}$ and $TR_{b,top}$ incorporate a temporal correction which accounts for whether said current field coded image is top field first or bottom field first.

8. The method of one of the preceding claims, wherein:

$MV_{b,top}$, the backward motion vector for predicting the top field of the current image is determined according to one of the equations (a) $MV_{b,top} = ((TR_{b,top}-TR_{b,top})*MV_{top})/TR_{b,top}$ and (b) $MV_{b,top} = MV_{f,top} - MV_{top}$;

where $TR_{b,top}$ corresponds to a temporal spacing between the top field of the current image and the field of the past image which is referenced by MV_{top} , $TR_{b,top}$ corresponds to a temporal spacing between the top field of the future image and the field of the past image which is referenced by MV_{top} , and $MV_{f,top}$ is the forward motion vector for predicting the top field of the current image.

9. The method of claim 8, wherein:

said equation (a) is selected when a delta motion vector $MV_b=0$, and said equation (b) is selected when $MV_b \neq 0$.

10. The method of one of the preceding claims, wherein:

$MV_{b,bot}$, the backward motion vector for predicting the bottom field of the current image is determined according to one of the equations (a) $MV_{b,bot} = ((TR_{b,bot} - TR_{b,bot}) * MV_{bot}) / TR_{b,bot}$ and (b) $MV_{b,bot} = MV_{f,bot} - MV_{bot}$;

where $TR_{b,bot}$ corresponds to a temporal spacing between the bottom field of the current image and the field of the past image which is referenced by MV_{bot} , $TR_{b,bot}$ corresponds to a temporal spacing between the bottom field of the future image and the field of the past image which is referenced by MV_{bot} , and $MV_{f,bot}$ is the forward motion vector for predicting the bottom field of the current image.

11. The method of claim 10, wherein:
said equation (a) is selected when a delta motion vector $MV_b=0$, and said equation (b) is selected when $MV_b \neq 0$.

12. A method for selecting a coding mode for a current predicted, field coded macroblock having top

and bottom fields, in a sequence of digital video images, comprising the steps of:

determining a forward sum of absolute differences error, $SAD_{forward, field}$ for the current macroblock relative to a past reference macroblock, which corresponds to a forward coding mode;

determining a backward sum of absolute differences error, $SAD_{backward, field}$ for the current macroblock relative to a future reference macroblock, which corresponds to a backward coding mode;

determining an average sum of absolute differences error, $SAD_{average, field}$ for the current macroblock relative to an average of said past and future reference macroblocks, which corresponds to an average coding mode; and

selecting said coding mode according to the minimum of said SADs.

13. The method of claim 12, comprising the further step of:

selecting said coding mode according to the minimum of respective sums of said SADs with corresponding bias terms which account for the number of required motion vectors of the respective coding modes.

14. The method of claim 12 or 13, wherein:

$SAD_{forward,field}$ is determined according to a sum of: (a) a sum of absolute differences for the top field of the current macroblock relative to a top field of the past reference macroblock, and (b) a sum of absolute differences for the bottom field of the current macroblock relative to a bottom field of the past reference macroblock.

15. The method of one of claims 12 to 14,
wherein:

$SAD_{backward,field}$ is determined according to a sum of: (a) a sum of absolute differences for the top field of the current macroblock relative to a top field of the future reference macroblock, and (b) a sum of absolute differences for the bottom field of the current macroblock relative to a bottom field of the future reference macroblock.

16. The method of one of claims 12 to 15,
wherein:

$SAD_{average,field}$ is determined according to a sum of: (a) a sum of absolute differences for the top field of the current macroblock relative to an average of the top fields of the past and future reference macroblocks, and (b) a sum of absolute differences for the bottom field of the current macroblock relative to an average of the bottom fields of the past and future reference macroblocks.

17. A decoder for recovering a current, direct mode, field coded macroblock having top and bottom fields in a sequence of digital video macroblocks from a received bitstream, wherein said current macroblock is bi-directionally predicted using a past field coded reference macroblock having top and bottom fields, and a future field coded reference macroblock having top and bottom fields, comprising:

means for recovering MV_{top} , a forward motion vector of the top field of the future macroblock which references one of the top and bottom fields of the past macroblock, and MV_{bot} , a forward motion vector of the bottom field of the future macroblock which references one of the top and bottom fields of said past macroblock; and

means for determining forward and backward motion vectors for predicting at least one of the top and bottom fields of the current macroblock by scaling the forward motion vector of the corresponding field of the future macroblock.

18. The decoder of claim 17, further comprising:

means for determining $MV_{s,top}$, the forward motion vector for predicting the top field of the current macroblock, according to the expression

$$(MV_{top} * TR_{B,top}) / TR_{D,top} + MV_D;$$

where $TR_{s,top}$ corresponds to a temporal spacing between the top field of the current macroblock and

the field of the past macroblock which is referenced by MV_{top} , $TR_{B,top}$ corresponds to a temporal spacing between the top field of the future macroblock and the field of the past macroblock which is referenced by MV_{top} , and MV_B is a delta motion vector.

19. The decoder of claim 18, wherein:

$MV_{f,top}$ is determined using integer division with truncation toward zero; and

MV_{top} and MV_{bot} are integer half-luma pel motion vectors.

20. The decoder of claim 18 or 19, wherein:

$TR_{B,top}$ and $TR_{D,top}$ incorporate a temporal correction which accounts for whether said current field coded image is top field first or bottom field first.

21. The decoder of one of claims 17 to 20, further comprising:

means for determining $MV_{f,bot}$, the forward motion vector for predicting the bottom field of the current macroblock, according to the expression
 $(MV_{bot} * TR_{B,bot}) / TR_{D,bot} + MV_B$;

where $TR_{B,bot}$ corresponds to a temporal spacing between the bottom field of the current macroblock and the field of the past macroblock which is referenced by MV_{bot} , $TR_{D,bot}$ corresponds to a temporal spacing between the bottom field of the future

macroblock and the field of the past macroblock which is referenced by MV_{bot} , and MV_b is a delta motion vector.

22. The decoder of claim 21, wherein:
 $MV_{f,bot}$ is determined using integer division with truncation toward zero; and
 MV_{top} and MV_{bot} are integer half-luma pel motion vectors.

23. The decoder of claim 21 or 22, wherein:
 $TR_{b,bot}$ and $TR_{b,top}$ incorporate a temporal correction which accounts for whether said current field coded image is top field first or bottom field first.

24. The decoder of one of claims 17 to 23, further comprising:

means for determining $MV_{b,top}$, the backward motion vector for predicting the top field of the current macroblock, according to one of the equations (a) $MV_{b,top} = ((TR_{b,top} - TR_{b,top}) * MV_{top}) / TR_{b,top}$ and (b) $MV_{b,top} = MV_{f,top} - MV_{top}$;

where $TR_{b,top}$ corresponds to a temporal spacing between the top field of the current macroblock and the field of the past macroblock which is referenced by MV_{top} , $TR_{b,top}$ corresponds to a temporal spacing between the top field of the future macroblock and the field of the past macroblock which is referenced

by MV_{top} , and $MV_{f,top}$ is the forward motion vector for predicting the top field of the current macroblock.

25. The decoder of claim 24, further comprising:

means for selecting said equation (a) when a delta motion vector $MV_d=0$; and

means for selecting said equation (b) when $MV_d \neq 0$.

26. The decoder of one of claims 17 to 24, further comprising:

means for determining $MV_{b,bet}$, the backward motion vector for predicting the bottom field of the current macroblock, according to one of the equations (a) $MV_{b,bet} = ((TR_{b,bet}-TR_{f,bet}) * MV_{bot}) / TR_{b,bet}$ and (b) $MV_{b,bet} = MV_{f,bet} - MV_{bot}$;

where $TR_{b,bet}$ corresponds to a temporal spacing between the bottom field of the current macroblock and the field of the past macroblock which is referenced by MV_{bot} , $TR_{f,bet}$ corresponds to a temporal spacing between the bottom field of the future macroblock and the field of the past macroblock which is referenced by MV_{bot} , and $MV_{f,bet}$ is the forward motion vector for predicting the bottom field of the current macroblock.

27. The decoder of claim 26, further comprising:

means for selecting said equation (a) when a delta motion vector $MV_d=0$; and

means for selecting said equation (b) when

$MV_d \neq C$.

3.Detailed Description of Invention

BACKGROUND OF THE INVENTION

5 The present invention provides a method and apparatus for coding of digital video images such as bi-directionally predicted video object planes (B-VOPs), in particular, where the B-VOP and/or a reference image used to code the B-VOP is interlaced coded.

10 The invention is particularly suitable for use with various multimedia applications, and is compatible with the MPEG-4 Verification Model (VM) 8.0 standard (MPEG-4 VM 8.0) described in document ISO/IEC/JTC1/SC29/WG11 N1796, entitled "MPEG-4 Video 15 Verification Model Version 8.0", Stockholm, July 1997, incorporated herein by reference. The MPEG-2 standard is a precursor to the MPEG-4 standard, and is described in document ISO/IEC 13818-2, entitled "Information Technology - Generic Coding of Moving 20 Pictures and Associated Audio, Recommendation H.262," March 25, 1994, incorporated herein by reference.

25 MPEG-4 is a coding standard which provides a flexible framework and an open set of coding tools for communication, access, and manipulation of digital audio-visual data. These tools support a

wide range of features. The flexible framework of MPEG-4 supports various combinations of coding tools and their corresponding functionalities for applications required by the computer,
5 telecommunication, and entertainment (i.e., TV and film) industries, such as database browsing, information retrieval, and interactive communications.

MPEG-4 provides standardized core technologies
10 allowing efficient storage, transmission and manipulation of video data in multimedia environments. MPEG-4 achieves efficient compression, object scalability, spatial and temporal scalability, and error resilience.

The MPEG-4 video VM coder/decoder (codec) is a
15 block- and object-based hybrid coder with motion compensation. Texture is encoded with an 8x8 Discrete Cosine Transformation (DCT) utilizing overlapped block-motion compensation. Object shapes
20 are represented as alpha maps and encoded using a Content-based Arithmetic Encoding (CAE) algorithm or a modified DCT coder, both using temporal prediction. The coder can handle sprites as they
25 are known from computer graphics. Other coding methods, such as wavelet and sprite coding, may also be used for special applications.

Motion compensated texture coding is a well known approach for video coding, and can be modeled as a three-stage process. The first stage is signal processing which includes motion estimation and
30 compensation (ME/MC) and a two-dimensional (2-D)

5 spatial transformation. The objective of ME/MC and the spatial transformation is to take advantage of temporal and spatial correlations in a video sequence to optimize the rate-distortion performance of quantization and entropy coding under a complexity constraint. The most common technique for ME/MC has been block matching, and the most common spatial transformation has been the DCT.

10 However, special concerns arise for ME/MC of macroblocks (MBs) in B-VOPs when the MB is itself interlaced coded and/or uses reference images which are interlaced coded.

15 In particular, it would be desirable to have an efficient technique for providing motion vector (MV) predictors for a MB in a B-VCP. It would also be desirable to have an efficient technique for direct mode coding of a field coded MB in a B-VCP. It would further be desirable to have a coding mode decision process for a MB in a field coded B-VOP for selecting the reference image which is results in the most efficient coding.

20 The present invention provides a system having the above and other advantages.

SUMMARY OF THE INVENTION

In accordance with the present invention, a method and apparatus are presented for coding of digital video images such as a current image (e.g., macroblock) in a bi-directionally predicted video object plane (B-VOP), in particular, where the current image and/or a reference image used to code the current image is interlaced (e.g., field) coded.

In a first aspect of the invention, a method provides direct mode motion vectors (MVs) for a current bi-directionally predicted, field coded image such as a macroblock (MB) having top and bottom fields, in a sequence of digital video images. A past field coded reference image having top and bottom fields, and a future field coded reference image having top and bottom fields are determined. The future image is predicted using the past image such that MV_{tcp} , a forward MV of the top field of the future image, references either the top or bottom field of said past image. The field which is referenced contains a best-match MB for a MB in the top field of the future image.

This MV is termed a "forward" MV since, although it references a past image (e.g., backward in time), the prediction is from the past image to the future image, e.g., forward in time. As a mnemonic, the prediction direction may be thought of as being opposite the direction of the corresponding MV.

Similarly, MV_{bot} , a forward motion vector of the bottom field of the future image, references either the top or bottom field of the past image. Forward and backward MVs are determined for predicting the top and/or bottom fields of the current image by scaling the forward MV of the corresponding field of the future image.

In particular, $MV_{f,top}$, the forward motion vector for predicting the top field of the current image, is determined according to the expression $MV_{f,top} = (MV_{top} * TR_{s,top}) / TR_{D,top} + MV_D$, where MV_D is a delta motion vector for a search area, $TR_{s,top}$ corresponds to a temporal spacing between the top field of the current image and the field of the past image which is referenced by MV_{top} , and $TR_{D,top}$ corresponds to a temporal spacing between the top field of the future image and the field of the past image which is referenced by MV_{top} . The temporal spacing may be related to a frame rate at which the images are displayed.

Similarly, $MV_{f,bot}$, the forward motion vector for predicting the bottom field of the current image, is determined according to the expression $MV_{f,bot} = (MV_{bot} * TR_{s,bot}) / TR_{D,bot} + MV_D$, where MV_D is a delta motion vector, $TR_{s,bot}$ corresponds to a temporal spacing between the bottom field of the current image and the

field of the past image which is referenced by MV_{bot} , and $TR_{D,bot}$ corresponds to a temporal spacing between the bottom field of the future MB and the field of the past MB which is referenced by MV_{bot} .

5 $MV_{b,top}$, the backward motion vector for predicting the top field of the current MB is determined according to the equation $MV_{b,top} = ((TR_{B,top} - TR_{E,top}) * MV_{top}) / TR_{D,top}$ when the delta motion vector $MV_D = 0$, or $MV_{b,top} = MV_{f,top} - MV_{top}$ when $MV_D \neq 0$.

10 $MV_{b,bot}$, the backward motion vector for predicting the bottom field of the current MB is determined according to the equation $MV_{b,bot} = ((TR_{B,bot} - TR_{E,bot}) * MV_{bot}) / TR_{D,bot}$ when the delta motion vector $MV_D = 0$, or $MV_{b,bot} = MV_{f,bot} - MV_{bot}$ when $MV_D \neq 0$.

15 A corresponding decoder is also presented. In another aspect of the invention, a method is presented for selecting a coding mode for a current predicted, field coded MB having top and bottom fields, in a sequence of digital video MBs. The coding mode may be a backward mode, where the reference MB is temporally after the current MB in display order, a forward mode, where the reference MB is before the current MB, or average (e.g., bidirectional) mode, where an average of prior and subsequent reference MBs is used.

The method includes the step of determining a forward sum of absolute differences error, SAD_{forward,field} for the current MB relative to a past reference MB, which corresponds to a forward coding mode. SAD_{forward,field} indicates the error in pixel luminance values between the current MB and a best match MB in the past reference MB. A backward sum of absolute differences error, SAD_{backward,field} for the current MB relative to a future reference MB, which corresponds to a backward coding mode is also determined. SAD_{backward,field} indicates the error in pixel luminance values between the current MB and a best match MB in the future reference MB.

An average sum of absolute differences error, SAD_{average,field} for the current MB relative to an average of the past and future reference MBs, which corresponds to an average coding mode, is also determined. SAD_{average,field} indicates the error in pixel luminance values between the current MB and a MB which is the average of the best match MBs of the past and future reference MBs.

The coding mode is selected according to the minimum of the SADs. Bias terms which account for the number of required MVs of the respective coding modes may also be factored into the coding mode selection process.

$SAD_{forward, field}$, $SAD_{backward, field}$, and $SAD_{average, field}$
are determined by summing the component terms over
the top and bottom fields.

DETAILED DESCRIPTION OF THE INVENTION

A method and apparatus are presented for coding of a digital video image such as a macroblock (MB) in a bi-directionally predicted video object plane (B-VOP), in particular, where the MB and/or a reference image used to code the MB is interlaced coded. The scheme provides a method for selecting a prediction motion vector (PMV) for the top and bottom field of a field coded current MB, including forward and backward PMVs as required, as well as for frame coded MBs. A direct coding mode for a field coded MB is also presented, in addition to a coding decision process which uses the minimum of sum of absolute differences terms to select an optimum mode.

FIG. 1 is an illustration of a video object plane (VOP) coding and decoding process in accordance with the present invention. Frame 105 includes three pictorial elements, including a square foreground element 107, an oblong foreground element 108, and a landscape backdrop element 109. In frame 115, the elements are designated VOPs using a segmentation mask such that VOP 117 represents the square foreground element 107, VOP 118 represents the oblong foreground element 108, and VOP 119 represents the landscape backdrop element 109. A VOP can have an arbitrary shape, and a succession of VOPs is known as a video object. A full rectangular video frame may also be considered to be a VOP.

Thus, the term "VOP" will be used herein to indicate both arbitrary and non-arbitrary (e.g., rectangular) image area shapes. A segmentation mask is obtained using known techniques, and has a format similar to that of ITU-R 601 luminance data. Each pixel is identified as belonging to a certain region in the video frame.

The frame 105 and VOP data from frame 115 are supplied to separate encoding functions. In particular, VOPs 117, 118 and 119 undergo shape, motion and texture encoding at encoders 137, 138 and 139, respectively. With shape coding, binary and gray scale shape information is encoded. With motion coding, the shape information is coded using motion estimation within a frame. With texture coding, a spatial transformation such as the DCT is performed to obtain transform coefficients which can be variable-length coded for compression.

The coded VOP data is then combined at a multiplexer (MUX) 140 for transmission over a channel 145. Alternatively, the data may be stored on a recording medium. The received coded VOP data is separated by a demultiplexer (DEMUX) 150 so that the separate VOPs 117-119 are decoded and recovered. Frames 155, 165 and 175 show that VOPs 117, 118 and 119, respectively, have been decoded and recovered and can therefore be individually manipulated using a compositor 160 which interfaces with a video library 170, for example.

The compositor may be a device such as a personal computer which is located at a user's home

to allow the user to edit the received data to provide a customized image. For example, the user's personal video library 170 may include a previously stored VOP 178 (e.g., a circle) which is different than the received VOPs. The user may compose a frame 185 where the circular VOP 178 replaces the square VOP 117. The frame 185 thus includes the received VOPs 118 and 119 and the locally stored VOP 178.

In another example, the background VOP 109 may be replaced by a background of the user's choosing. For example, when viewing a television news broadcast, the announcer may be coded as a VOP which is separate from the background, such as a news studio. The user may select a background from the library 170 or from another television program, such as a channel with stock price or weather information. The user can therefore act as a video editor.

The video library 170 may also store VOPs which are received via the channel 145, and may access VOPs and other image elements via a network such as the Internet. Generally, a video session comprises a single VOP, or a sequence of VOPs.

The video object coding and decoding process of FIG. 1 enables many entertainment, business and educational applications, including personal computer games, virtual environments, graphical user interfaces, videoconferencing, Internet applications and the like. In particular, the capability for ME/MC with interlaced coded (e.g., field mode) VOPs

in accordance with the present invention provides even greater capabilities.

FIG. 2 is a block diagram of an encoder in accordance with the present invention. The encoder is suitable for use with both predictive-coded VOPs (P-VOPs) and bi-directionally coded VOPs (B-VOPs).

P-VOPs may include a number of macroblocks (MBs) which may be coded individually using an intra-frame mode or an inter-frame mode. With intra-frame (INTRA) coding, the macroblock (MB) is coded without reference to another MB. With inter-frame (INTER) coding, the MB is differentially coded with respect to a temporally subsequent frame in a mode known as forward prediction. The temporally subsequent frame is known as an anchor frame or reference frame. The anchor frame (e.g., VOP) must be a P-VOP or an I-VOP, not a B-VOP. An I-VOP includes self-contained (e.g., intra-coded) blocks which are not predictive coded.

With forward prediction, the current MB is compared to a search area of MBs in the anchor frame to determine the best match. A corresponding motion vector (MV), known as a backward MV, describes the displacement of the current MB relative to the best match MB. Additionally, an advanced prediction mode for P-VOPs may be used, where motion compensation is performed on 8x8 blocks rather than 16x16 MBs. Moreover, both intra-frame and inter-frame coded P-VOP MBs can be coded in a frame mode or a field mode.

B-VOPs can use the forward prediction mode as described above in connection with P-VOPs as well as backward prediction, bi-directional prediction, and direct mode, which are all inter-frame techniques.

5 B-VOPs do not currently use intra-frame coded MBs under MPEG-4 VM 8.0, although this is subject to change. The anchor frame (e.g., VOP) must be a P-VOP or I-VOP, not a B-VOP.

With backward prediction of B-VOPs, the current
10 MB is compared to a search area of MBs in a temporally previous anchor frame to determine the best match. A corresponding MV, known as a forward MV), describes the relative displacement of the current MB relative to the best match MB. With bi-directional prediction of a B-VOP MB, the current MB
15 is compared to a search area of MBs in both a temporally previous anchor frame and a temporally subsequent anchor frame to determine the best match MBs. Forward and backward MVs describe the
20 displacement of the current MB relative to the best match MBs. Additionally, an averaged image is obtained from the best match MBs for use in encoding the current MB.

With direct mode prediction of B-VOPs, a MV is
25 derived for an 8x8 block when the collocated MB in the following P-VOP uses the 8x8 advanced prediction mode. The MV of the 8x8 block in the P-VOP is linearly scaled to derive a MV for the block in the B-VOP without the need for searching to find a best
30 match block.

The encoder, shown generally at 200, includes a shape coder 210, a motion estimation function 220, a motion compensation function 230, and a texture coder 240, which each receive video pixel data input at terminal 205. The motion estimation function 220, motion compensation function 230, texture coder 240, and shape coder 210 also receive VOP shape information input at terminal 207, such as the MPEG-4 parameter `VOP_of_arbitrary_shape`. When this parameter is zero, the VOP has a rectangular shape, and the shape coder 210 therefore is not used.

A reconstructed anchor VOP function 250 provides a reconstructed anchor VOP for use by the motion estimation function 220 and motion compensation function 230. A current VOP is subtracted from a motion compensated previous VOP at subtractor 260 to provide a residue which is encoded at the texture coder 240. The texture coder 240 performs the DCT to provide texture information (e.g., transform coefficients) to a multiplexer (MUX) 280. The texture coder 240 also provides information which is summed with the output from the motion compensator 230 at a summer 270 for input to the previous reconstructed VOP function 250.

Motion information (e.g., motion vectors) is provided from the motion estimation function 220 to the MUX 280, while shape information which indicates the shape of the VOP is provided from the shape coding function 210 to the MUX 280. The MUX 280 provides a corresponding multiplexed data stream to

a buffer 290 for subsequent communication over a data channel.

The pixel data which is input to the encoder may have a YUV 4:2:0 format. The VOP is represented by means of a bounding rectangle. The top left coordinate of the bounding rectangle is rounded to the nearest even number not greater than the top left coordinates of the tightest rectangle. Accordingly, the top left coordinate of the bounding rectangle in the chrominance component is one-half that of the luminance component.

FIG. 3 illustrates an interpolation scheme for a half-pixel search. Motion estimation and motion compensation (ME/MC) generally involve matching a block of a current video frame (e.g., a current block) with a block in a search area of a reference frame (e.g., a predicted block or reference block). For predictive (P) coded images, the reference block is in a previous frame. For bi-directionally predicted (B) coded images, predicted blocks in previous and subsequent frames may be used. The displacement of the predicted block relative to the current block is the motion vector (MV), which has horizontal (x) and vertical (y) components. Positive values of the MV components indicate that the predicted block is to the right of, and below, the current block.

A motion compensated difference block is formed by subtracting the pixel values of the predicted block from those of the current block point by point. Texture coding is then performed on the

difference block. The coded MV and the coded texture information of the difference block are transmitted to the decoder. The decoder can then reconstruct an approximated current block by adding the quantized difference block to the predicted block according to the MV. The block for ME/MC can be a 16x16 frame block (macroblock), an 8x8 block or a 16x8 field block.

Accuracy of the MV is set at half-pixel.
Interpolation must be used on the anchor frame so that $p(i+x,j+y)$ is defined for x or y being half of an integer. Interpolation is performed as shown in FIG. 3. Integer pixel positions are represented by the symbol "+", as shown at A, B, C and D. Half-pixel positions are indicated by circles, as shown at a, b, c and d. As seen, $a = A$, $b = (A + B) // 2$, $c = (A + C) // 2$, and $d = (A + B + C + D) // 4$, where "://" denotes rounded division. Further details of the interpolation are discussed in MPEG-4 VM 8.0 referred to previously as well as commonly assigned U.S. Patent application Serial No. 08/897,847 to Eifrig et al., filed July 21, 1997, entitled "Motion Estimation and Compensation of Video Object Planes for Interlaced Digital Video", incorporated herein by reference.

FIG. 6 illustrates reordering of pixel lines in an adaptive frame/field prediction scheme in accordance with the present invention. In a first aspect of the advanced prediction technique, an adaptive technique is used to decide whether a

current macroblock (MB) of 16x16 pixels should be
ME/MC coded as is, or divided into four blocks of
8x8 pixels each, where each 8x8 block is ME/MC coded
separately, or whether field based motion estimation
should be used, where pixel lines of the MB are
reordered to group the same-field lines in two 16x8
field blocks, and each 16x8 block is separately
ME/MC coded.

A field mode 16x16 macroblock (MB), is shown
generally at 600. The MB includes even-numbered
lines 602, 604, 606, 608, 610, 612, 614 and 616, and
odd-numbered lines 603, 605, 607, 609, 611, 613, 615
and 617. The even and odd lines are thus
interleaved, and form top and bottom (or first and
second) fields, respectively.

When the pixel lines in image 600 are permuted
to form same-field luminance blocks, the MB shown
generally at 650 is formed. Arrows, shown generally
at 645, indicate the reordering of the lines 602-
617. For example, the even line 602, which is the
first line of MB 600, is also the first line of MB
650. The even line 604 is reordered as the second
line in MB 650. Similarly, the even lines 606, 608,
610, 612, 614 and 616 are reordered as the third
through eighth lines, respectively, of MB 650.
Thus, a 16x8 luminance region 680 with even-numbered
lines is formed. Similarly, the odd-numbered lines
603, 605, 607, 609, 611, 613, 615 and 617 form a
16x8 region 685.

The decision process for choosing the MC mode
for P-VOPs is as follows. For frame mode video,

first obtain the Sum of Absolute Differences (SAD) for a single 16x16 MB, e.g., $SAD_{16}(MV_x, MV_y)$; and for four 8x8 blocks, e.g.,

$SAD_4(MV_{x1}, MV_{y1}), SAD_4(MV_{x2}, MV_{y2}), SAD_4(MV_{x3}, MV_{y3}),$ and $SAD_4(MV_{x4}, MV_{y4})$.

- 5 If $\sum_{i=1}^4 SAD_4(MV_{xi}, MV_{yi}) < SAD_{16}(MV_x, MV_y) - 129$, choose 8x8 prediction; otherwise, choose 16x16 prediction. The constant "129" is obtained from $Nb/2+1$, where Nb is the number of non-transparent pixels in a MB.

For interlaced video, obtain

- 10 $SAD_{top}(MV_{x_top}, MV_{y_top}), SAD_{bottom}(MV_{x_bottom}, MV_{y_bottom})$, where (MV_{x_top}, MV_{y_top}) and $(MV_{x_bottom}, MV_{y_bottom})$ are the motion vector (MV) for both top (even) and bottom (odd) fields. Then, choose the reference field which has the smallest SAD (e.g., for SAD_{top} and SAD_{bottom}) from the field half sample search.

15 The overall prediction mode decision is based on choosing the minimum of:

$$(a) SAD_{16}(MV_x, MV_y), \quad (b) \sum_{i=1}^4 SAD_4(MV_{xi}, MV_{yi}) + 129,$$

and (c) $SAD_{top}(MV_{x_top}, MV_{y_top}) + SAD_{bottom}(MV_{x_bottom}, MV_{y_bottom}) + 65$.

- 20 If term (a) is the minimum, 16x16 prediction is used. If term (b) is the minimum, 8x8 motion compensation (advanced prediction mode) is used. If term (c) is the minimum, field based motion estimation is used. The constant "65" is obtained from $Nb/4+1$.

25 If 8x8 prediction is chosen, there are four MVs for the four 8x8 luminance blocks, i.e., one MV for

each 8x8 block. The MV for the two chrominance blocks is then obtained by taking an average of these four MVS and dividing the average value by two. Since each MV for the 8x8 luminance block has a half-pixel accuracy, the MV for the chrominance blocks may have a sixteenth pixel value. Table 1, below, specifies the conversion of a sixteenth pixel value to a half-pixel value for chrominance MVs. For example, 0 through 2/16 are rounded to 0, 3/16 through 13/16 are rounded to 1/2, and 14/16 and 15/16 are rounded to 2/2=1.

Table 1

1/16 pixel value	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
1/2 pixel value	0	0	0	1	1	1	1	1	1	1	1	1	1	1	2	2

With field prediction, there are two MVs for the two 16x8 blocks. The luminance prediction is generated as follows. The even lines of the MB (e.g., lines 602, 604, 606, 608, 610, 612, 614 and 616) are defined by the top field MV using the reference field specified. The MV is specified in frame coordinates such that full pixel vertical displacements correspond to even integral values of the vertical MV coordinate, and a half-pixel vertical displacement is denoted by odd integral values. When a half-pixel vertical offset is

specified, only pixels from lines within the same reference field are combined.

5 The MV for the two chrominance blocks is derived from the (luminance) MV by dividing each component by two, then rounding. The horizontal component is rounded by mapping all fractional values into a half-pixel offset. The vertical MV component is an integer and the resulting chrominance MV vertical component is rounded to an 10 integer. If the result of dividing by two yields a non-integral value, it is rounded to the adjacent odd integer. Note that the odd integral values denote vertical interpolation between lines of the same field.

15 The second aspect of the advanced prediction technique is overlapped MC for luminance blocks, discussed in greater detail in MPEG-4 VM 8.0 and Eifrig et al. application referred to previously.

20 Specific coding techniques for B-VOPs are now discussed. For INTER coded VOPs such as B-VOPs, there are four prediction modes, namely, direct mode, interpolate (e.g., averaged or bi-directional) mode, backward mode, and forward mode. The latter three modes are non-direct modes. Forward only, or 25 backward only prediction are also known as "unidirectional" prediction. The predicted blocks of the B-VOP are determined differently for each mode. Furthermore, blocks of a B-VOP and the anchor block(s) may be progressive (e.g., frame) coded or 30 interlaced (e.g., field) coded.

A single B-VOP can have different MBs which are predicted with different modes. The term "B-VOP" only indicates that bi-directionally predicted blocks may be included, but this is not required.

5 In contrast, with P-VOPs and I-VOPs, bi-directionally predicted MBs are not used.

For non-direct mode B-VOP MBs, MVs are coded differentially. For forward MVs in forward and bi-directional modes, and backward MVs in backward and bi-directional modes, the "same-type" MV (e.g., forward or backward) of the MB which immediately precedes the current MB in the same row is used as a predictor. This is the same as the immediately preceding MB in raster order, and generally, in transmission order. However, if the raster order differs from the transmission order, the MVs of the immediately preceding MB in transmission order should be used to avoid the need to store and re-order the MBs and corresponding MVs at the decoder.

20 Using the same-type MV, and assuming the transmission order is the same as the raster order, and that the raster order is from left to right, top to bottom, the forward MV of the left-neighboring MB is used as a predictor for the forward MV of the current MB of the B-VOP. Similarly, the backward MV of the left-neighboring MB is used as a predictor for the backward MV of the current MB of the B-VOP. The MVs of the current MB are then differentially encoded using the predictors. That is, the 25 difference between the predictor and the MV which is determined for the current MB is transmitted as a

motion vector difference to a decoder. At the decoder, the MV of the current MB is determined by recovering and adding the PMV and the difference MV.

In case the current MB is located on the left edge of the VOP, the predictor for the current MB is set to zero.

For interlaced-coded B-VOPs, each of the top and bottom fields have two associated prediction motion vectors, for a total of four MVs.

10 The four prediction MVs represent, in transmission order, the top field forward and bottom field forward of the previous anchor MB, and the top field backward and bottom field backward of the next anchor MB. The current MB and the forward MB,
15 and/or the current MB and the backward MB, may be separated by one or more intermediate images which are not used for ME/MC coding of the current MB. B-VOPs do not contain INTRA coded MBs, so each MB in the B-VOP will be ME/MC coded. The forward and backward anchor MBs may be from a P-VOP or I-VOP,
20 and may be frame or field coded.

For interlaced, non-direct mode B-VOP MBs, four possible prediction motion vectors (PMVs) are shown in Table 2 below. The first column of Table 2 shows the prediction function, while the second column shows a designator for the PMV. These PMVs are used as shown in Table 3 below for the different MB prediction modes.

33

Table 2

Prediction function	PMV type
Top field, forward	0
Bottom field, forward	1
Top field, backward	2
Bottom field, backward	3

Table 3

Macroblock mode	PMV type used
Frame, forward	0
Frame, backward	2
Frame, bi-directional	0,2
Field, forward	0,-
Field, backward	2,3
Field, bi-directional	0,-,2,3

For example, Table 3 shows that, for a current field mode MB with a forward prediction mode (e.g., "Field, forward"), top field forward ("0") and bottom field forward ("1") motion vector predictors are used.

After being used in differential coding, the motion vectors of a current MB become the PMVs for a subsequent MB, in transmission order. The PMVs are reset to zero at the beginning of each row of MBs since the MVs of a MB at the end of a preceding row are unlikely to be similar to the MVs of a MB at the beginning of a current row. The predictors are also not used for direct mode MBs. For skipped MBs, the PMVs retain the last value.

With direct mode coding of B-VOP MBs, no vector differences are transmitted. Instead, the forward and backward MVs are directly computed at the

decoder from the MVs of the temporally next P-VOP MB, with correction by a single delta MV, which is not predicted. The technique is efficient since less MV data is transmitted.

5 Table 4 below summarizes which PMVs are used to code the motion vectors of the current B-VOP MB based on the previous and current MB types. For B-VOPs, an array of prediction motion vectors, pmv[] may be provided which are indexed from zero to three
10 (e.g., pmv[0], pmv[1], pmv[2] and pmv[3]). The indexes pmv[] are not transmitted, but the decoder can determine the pmv[] index to use according to the MV coding type and the particular vector being decoded. After coding a B-VOP MB, some of the PMVs
15 vectors are updated to be the same as the motion vectors of the current MB. The first one, two or four PMVs are updated depending on the number of MVs associated with the current MB.

20 For example, a forward, field predicted MB has two motion vectors, where pmv[0] is the PMV for the top field, forward, and pmv[1] is the PMV for the bottom field, forward. For a backward, field predicted MB, pmv[2] is the PMV for the top field backward, and pmv[3] is the PMV for the bottom field backward. For a bi-directional, field predicted MB, pmv[0] is the PMV for the top field, forward, pmv[1] is the PMV for the bottom field forward, pmv[2] is the PMV for the top field backward, and pmv[3] is the PMV for the bottom field backward. For a forward or backward predicted frame mode B-VOP MB, there is only one MV, so only pmv[0]

is used for forward, and pmv[2] is used for backward. For an average (e.g., bi-directionally) predicted frame mode B-VOP MB, there are two MVs, namely, pmv[0] for the forward MV, and pmv[2] for the backward MV. The row designated "pmv[]'s to update" indicates whether one, two or four MVs are updated.

{}

Table 4 - Prediction Motion Vector Index pmv[]

Previous Macroblock type in transmission order	Current Macroblock type			
	Direct	Forward, Frame Mode	Backward, Frame Mode	Average, Frame Mode
pmv[]'s to update	none	0,1	2,3	0,1,2,3
pmv[]'s to use	none	0	2	0,2
				0,1
				2,3
				2,3
				0,1,2,3

It will be appreciated that Table 4 is merely a shorthand notation for implementing the technique of the present invention for selecting a prediction MV for a current MB. However, the scheme may be expressed in various other ways.

Intra block DC adaptive prediction can use the same algorithm as described in MPEG-4 VM 8.0 regardless of value of `dct_type`. Intra block adaptive AC prediction is performed as described in MPEG-4 VM 8.0 except when the first `rcw` of coefficients is to be copied from the coded block above. This operation is allowed only if `dct_type` has the same value for the current MB and the block above. If the `dct_types` differ, then AC prediction can occur only by copying the first column from the block to the left. If there is no left block, zero is used for the AC predictors.

FIG. 4 illustrates direct mode coding of the top field of an interlaced-coded B-VOP in accordance with the present invention. Progressive direct coding mode is used for the current macroblock (MB) whenever the MB in a future anchor picture which is at the same relative position (e.g., co-sited) as the current MB is coded as (1) a 16x16 (frame) MB, (2) an intra MB or (3) an 8x8 (advanced prediction) MB.

The direct mode prediction is interlaced whenever the co-sited future anchor picture MB is coded as an interlaced MB. Direct mode will be used to code the current MB if its biased SAD is the

} A

minimum of all B-VOP MB predictors. Direct mode for
an interlaced coded MB forms the prediction MB
separately for the top and bottom fields of the
current MB. The four field motion vectors (MVs) of
5 a bi-directional field motion compensated MB (e.g.,
top field forward, bottom field forward, top field
backward, and bottom field backward) are calculated
directly from the respective MVs of the
corresponding MB of the future anchor picture.

10 The technique is efficient since the required
searching is significantly reduced, and the amount
of transmitted MV data is reduced. Once the MVs and
reference field are determined, the current MB is
considered to be a bi-directional field predicted
15 MB. Only one delta motion vector (used for both
fields) occurs in the bitstream for the field
predicted MB.

20 The prediction for the top field of the current
MB is based on the top field MV of the MB of the
future anchor picture (which can be a P-VOP, or an
I-VOP with MV=0), and a past reference field of a
previous anchor picture which is selected by the
corresponding MV of the top field of the future
25 anchor MB. That is, the top field MB of the future
anchor picture which is correspondingly positioned
(e.g., co-sited) to the current MB has a best match
MB in either the top or bottom field of the past
anchor picture. This best match MB is then used as
the anchor MB for the top field of the current MB.
30 An exhaustive search is used to determine the delta

motion vector MV, given the co-sited future anchor MV on a MB by MB basis.

5 Motion vectors for the bottom field of the current MB are similarly determined using the MV of the correspondingly positioned bottom field of the future anchor MB, which in turn references a best match MB in the top or bottom field of the past anchor picture.

10 Essentially, the top field motion vector is used to construct an MB predictor which is the average of (a) pixels obtained from the top field of the correspondingly positioned future anchor MB and (b) pixels from the past anchor field referenced by the top field MV of the correspondingly positioned future anchor MB. Similarly, the bottom field motion vector is used to construct a MB predictor which is the average of (a) pixels obtained from the bottom field of the correspondingly positioned future anchor MB and (b) pixels from the past anchor field referenced by the bottom field MV of the correspondingly positioned future anchor MB.

15 20 25 As shown in FIG. 4, the current B-VOP MB 420 includes a top field 430 and bottom field 425, the past anchor VOP MB 400 includes a top field 410 and bottom field 405, and the future anchor VOP MB 440 includes a top field 450 and bottom field 445.

30 The motion vector MV_{top} is the forward motion vector for the top field 450 of the future anchor MB 440 which indicates the best match MB in the past anchor MB 400. Even though MV_{top} is referencing a

previous image (e.g., backward in time), it is a forward MV since the future anchor VOP 440 is forward in time relative to the past anchor VOP 400. In the example, MV_{top} references the bottom field 405 of the past anchor MB 400, although either the top 410 or bottom 405 field could be referenced. $MV_{f,top}$ is the forward MV of the top field of the current MB, and $MV_{b,top}$ is the backward MV of the top field of the current MB. Pixel data is derived for the bi-directionally predicted MB at a decoder by averaging the pixel data in the future and past anchor images which are identified by $MV_{b,top}$ and $MV_{f,top}$, respectively, and summing the averaged image with a residue which was transmitted.

The motion vectors for the top field are calculated as follows:

$$MV_{f,top} = (TR_{B,top} * MV_{top}) / TR_{D,top} + MV_D;$$

$$MV_{b,top} = ((TR_{B,top} - TR_{D,top}) * MV_{top}) / TR_{D,i}$$

if $MV_D=0$; and

$$MV_{b,top} = (MV_{f,top} - MV_{top}) \text{ if } MV_D \neq 0.$$

MV_D is a delta, or offset, motion vector. Note that the motion vectors are two-dimensional. Additionally, the motion vectors are integral half-pixel luma motion vectors. The slash "/" denotes truncate toward zero integer division. Also, the future anchor VOP is always a P-VOP for field direct

mode. If the future anchor was an I-VOP, the MV would be zero and 16x16 progressive direct mode would be used. $TR_{S,top}$ is the temporal distance in fields between the past reference field (e.g., top or bottom), which is the bottom field 405 in this example, and the top field 430 of the current B-VOP.

5 420. $TR_{D,top}$ is the temporal distance between the past reference field (e.g., top or bottom), which is the bottom field 405 in this example, and the future top reference field 450.

10 FIG. 5 illustrates direct mode coding of the bottom field of an interlaced-coded B-VOP in accordance with the present invention. Note that the source interlaced video can have a top field first or bottom field first format. A bottom field first format is shown in FIGS 4 and 5. Like-numbered elements are the same as in FIG. 4. Here, the motion vector MV_{bot} is the forward motion vector for the bottom field 445 of the future anchor macroblock (MB) 440 which indicates the best match MB in the past anchor MB 400. In the example, MV_{bot} references the bottom field 405 of the past anchor MB 400, although either the top 410 or bottom 405 field could be used. $MV_{f,bot}$ and $MV_{b,bot}$ are the forward and backward motion vectors, respectively.

15 20 25

The motion vectors for the bottom field are calculated in a parallel manner to the top field motion vectors, as follows:

```

    MVf,bot = (TRB,bot*MVbot) / TRD,bot+MVD;
    MVb,bot = ((TRB,bot - TRD,bot) * MVbot) / TRD,i
    if MVD=0; and
    MVb,bot = (MVf,bot - MVbot) if MVD≠0.

```

5 TR_{B,bot} is the temporal distance between the past reference field (e.g., top or bottom), which is the bottom field 405 in this example, and the bottom field 425 of the current B-VOP 420. TR_{D,bot} is the temporal distance between the past reference field 10 (e.g., top or bottom), which is the bottom field 405 in this example, and the future bottom reference field 445.

Regarding the examples of FIGS 4 and 5, the calculation of TR_{s,top}, TR_{D,top}, TR_{B,bot} and TR_{C,bot} depends not only on the current field, reference field, and frame temporal references, but also on whether the current video is top field first or bottom field first. In particular,

20 TR_{D,top} or TR_{D,bot} = 2*(TR_{future} - TR_{past}) + δ; and
TR_{B,top} or TR_{B,bot} = 2*(TR_{current} - TR_{past}) + δ;

where TR_{future}, TR_{current}, and TR_{past} are the frame number of the future, current and past frames, respectively, in display order, and δ, an additive correction to the temporal distance between fields, 25 is given by Table 5, below. δ has units of field periods.

43

For example, the designation "1" in the last row of the first column indicates that the future anchor field is the top field, and the referenced field is the bottom field. This is shown in FIG. 4.

5 The designation "1" in the last row of the second column indicates that the future anchor field is the bottom field, and the referenced field is also the bottom field. This is shown in FIG. 5.

Table 5 - Temporal correction, δ

Referenced Field		Bottom Field First		Top Field First	
Future Anchor= top	Future Anchor= bottom	Top Field δ	Bottom Field δ	Top Field δ	Bottom Field δ
top	top	0	-1	0	1
top	bottom	0	0	0	0
bottom	top	1	-1	-1	1
bottom	bottom	1	0	-1	0

10 For efficient coding, an appropriate coding mode decision process is required. As indicated, for B-VOPs, a MB can be coded using (1) direct coding, (2) 16x16 motion compensated (includes forward, backward and averaged modes), or (3) field motion compensation (includes forward, backward and averaged modes). Frame or field direct coding of a current MB is used when the corresponding future anchor MB is frame or field direct coded, respectively.

15 20 For a field motion compensated MB in a B-VOPs, a decision is made to code the MB in a forward,

backward, or averaged mode based on the minimum luminance half-pixel SADs with respect to the decoded anchor pictures. Specifically, seven biased SAD terms are calculated as follows:

5

- (1) $SAD_{direct} + b_1$,
- (2) $SAD_{forward} + b_2$,
- (3) $SAD_{backward} + b_2$,
- (4) $SAD_{average} - b_3$,
- (5) $SAD_{forward, field} + b_3$,
- (6) $SAD_{backward, field} + b_3$, and
- (7) $SAD_{average, field} + b_4$.

10 where the subscripts indicate direct mode, forward motion prediction, backward motion prediction, average (i.e., interpolated or bi-directional) motion prediction, frame mode (i.e., locally progressive) and field mode (i.e., locally interlaced). The field SADs abcve (i.e.,

15 $SAD_{forward, field}$, $SAD_{backward, field}$, and $SAD_{average, field}$) are the sums of the top and bottom field SADs, each with its own reference field and motion vector.

20 Specifically, $SAD_{forward, field} = SAD_{forward, top field} + SAD_{forward, bottom field}$; $SAD_{backward, field} = SAD_{backward, top field} + SAD_{backward, bottom field}$; and $SAD_{average, field} = SAD_{average, top field} + SAD_{average, bottom field}$.

25 SAD_{direct} is the best direct mode prediction, $SAD_{forward}$ is the best 16x16 prediction from the forward (past) reference, $SAD_{backward}$ is the best 16x16 prediction from the backward (future)

reference, $SAD_{average}$ is the best 16x16 prediction formed by a pixel-by-pixel average of the best forward and best backward reference, $SAD_{forward,field}$ is the best field prediction from the forward (past) reference, $SAD_{backward,field}$ is the best field prediction from the backward (future) reference, and $SAD_{average,field}$ is the best field prediction formed by a pixel-by-pixel average of the best forward and best backward reference.

The b_i 's are bias values as defined in Table 6, below, to account for prediction modes which require more motion vectors. Direct mode and modes with fewer MVs are favored.

Table 6

Mode	Number of motion vectors	b_i	Bias	Value
Direct	1	b_1	$-(Nb/2 + 1)$	-129
Frame, forward	1	b_2	0	0
Frame, backward	1	b_2	0	0
Frame, average	2	b_3	$(Nb/4 + 1)$	65
Field, forward	2	b_3	$(Nb/4 + 1)$	65
Field, backward	2	b_3	$(Nb/4 + 1)$	65
Field, average	4	b_4	$(Nb/2 + 1)$	129

The negative bias for direct mode is for consistency with the existing MPEG-4 VM for progressive video, and may result in relatively more skipped MBs.

FIG. 7 is a block diagram of a decoder in accordance with the present invention. The decoder, shown generally at 700, can be used to receive and decode the encoded data signals transmitted from the encoder of FIG. 2. The encoded video image data and differentially encoded motion vector (MV) data are received at terminal 740 and provided to a demultiplexer (DEMUX) 742. The encoded video image data is typically differentially encoded in DCT transform coefficients as a prediction error signal (e.g., residue).

A shape decoding function 744 processes the data when the VOP has an arbitrary shape to recover shape information, which is, in turn, provided to a motion compensation function 750 and a VOP reconstruction function 752. A texture decoding function 746 performs an inverse DCT on transform coefficients to recover residue information. For INTRA coded macroblocks (MBs), pixel information is recovered directly and provided to the VOP reconstruction function 752.

For INTER coded blocks and MBs, such as those in B-VOPs, the pixel information provided from the texture decoding function 746 to the reconstructed VOP function 752 represents a residue between the current MB and a reference image. The reference image may be pixel data from a single anchor MB which is indicated by a forward or backward MV.

Alternatively, for an interpolated (e.g., averaged) MB, the reference image is an average of pixel data from two reference MBs, e.g., one past anchor MB and one future anchor MB. In this case, the decoder must calculate the averaged pixel data according to the forward and backward MVs before recovering the current MB pixel data.

For INTER coded blocks and MBs, a motion decoding function 748 processes the encoded MV data to recover the differential MVs and provide them to the motion compensation function 750 and to a motion vector memory 749, such as a RAM. The motion compensation function 750 receives the differential MV data and determines a reference motion vector (e.g., predictor motion vector, or PMV) in accordance with the present invention. The PMV is determined according to the coding mode (e.g., forward, backward, bi-directional, or direct).

Once the motion compensation function 750 determines a full reference MV and sums it with the differential MV of the current MB, the full MV of the current MB is available. Accordingly, the motion compensation function 750 can now retrieve anchor frame best match data from a VOP memory 754, such as a RAM, calculate an averaged image if required, and provide the anchor frame pixel data to the VOP reconstruction function to reconstruct the current MB.

The retrieved or calculated best match data is added back to the pixel residue at the VOP reconstruction function 752 to obtain the decoded

49

current MB or block. The reconstructed block is output as a video output signal and also provided to the VOP memory 754 to provide new anchor frame data. Note that an appropriate video data buffering 5 capability may be required depending on the frame transmission and presentation orders since an anchor frame for a B-VOP MB may be a temporally future frame or field, in presentation order.

FIG. 8 illustrates a MB packet structure in accordance with the present invention. The structure is suitable for B-VOPS, and indicates the format of data received by the decoder. Note that the packets are shown in four rows for convenience only. The packets are actually transmitted 10 serially, starting from the top row, and from left to right within a row. The first row 810 includes fields first_shape_code, MVD_shape, CR, ST and BAC. A second row 830 includes fields MODE and MBTYPE. A third row 850 includes fields CBPB, DQUANT, 15 Interlaced_information, MVD_t, MVD_b, and MVDB. A fourth row includes fields CODA, CBPBA, Alpha Block Data and Block Data. Each of the above fields is defined according to MPEG-4 VM 8.0.

first_shape_code indicates whether a MB is in a 20 bounding box of a VOP. CR indicates a conversion ratio for Binary Alpha Blocks. ST indicates a horizontal or vertical scan order. BAC refers to a binary arithmetic codeword.

MODE, which indicates the mode of a MB, is 25 present for every coded (non-skipped) MB in a B-VOP. Difference motion vectors (MVD_t, MVD_b, or MVDB) and

CBPB are present if indicated by MODB. Macroblock type is indicated by MBTYPE, which also signals motion vector modes (MVDs) and quantization (DQUANT). With interlaced mode, there can be up to four MVs per MB. MBTYPE indicates the coding type, e.g., forward, backward, bi-directional or direct. CBPB is the Coded Block Pattern for a B-type macroblock. CBPBA is similarly defined as CBPB except that it has a maximum of four bits. DQUANT defines changes in the value of a quantizer.

The field Interlaced_information in the third row 850 indicates whether a MB is interlaced coded, and provides field MV reference data which informs the decoder of the coding mode of the current MB or block. The decoder uses this information in calculating the MV for a current MB. The Interlaced_information field may be stored for subsequent use as required in the MV memory 749 or other memory in the decoder.

The Interlaced_information field may also include a flag dct_type which indicates whether top and bottom field pixel lines in a field coded MB are reordered from the interleaved order, as discussed above in connection with FIG. 6.

The MB layer structure shown is used when VOP_prediction_type== 10. If COD indicates skipped (COD == "1") for a MB in the most recently decoded I- or P-VOP then the co-located (e.g., co-sited) MB in the B-VOP is also skipped. That is, no information is included in the bitstream.

MVD_e is the motion vector of a MB in B-VOP with respect to a temporally previous reference VOP (an I- or a P-VOP). It consists of a variable length codeword for the horizontal component followed by a variable length codeword for the vertical component. For an interlaced B-VOP MB with field_prediction of "1" and MBTYPE of forward or interpolate, MVD_e represents a pair of field motion vectors (top field followed by bottom field) which reference the past anchor VOP.

MVD_s is the motion vector of a MB in B-VOP with respect to temporally following reference VOP (an I- or a P-VOP). It consists of a variable length codeword for the horizontal component followed by a variable length codeword for the vertical component. For an interlaced B-VOP MB with field_prediction of "1" and MBTYPE of backward or interpolate, MVD_s represents a pair of field MVs (top field followed by bottom field) which reference the future anchor VOP.

MVDB is only present in B-VOPs if direct mode is indicated by MODE and MBTYPE, and consists of a variable length codeword for the horizontal component followed by a variable length codeword for the vertical component of each vector. MVDBs represents delta vectors that are used to correct B-VOP MB motion vectors which are obtained by scaling P-VOP MB motion vectors.

CODA refers to gray scale shape coding.
The arrangement shown in FIG. 8 is an example only and that various other arrangements for

communicating the relevant information to the decoder will become apparent to those skilled in the art.

5 A bitstream syntax and MB layer syntax for use in accordance with the present invention is described in MPEG-4 VM 8.0 as well as the Eifrig et al. application referred to previously.

10 Accordingly, it can be seen that the present invention provides a scheme for encoding a current MB in a B-VOP, in particular, when the current MB is field coded, and/or an anchor MB is field coded. A scheme for direct coding for a field coded MB is presented, in addition to a coding decision process which uses the minimum of sum of absolute 15 differences terms to select an optimum mode. A prediction motion vector (PMV) is also provided for the top and bottom field of a field coded current MB, including forward and backward PMVs as required, as well as for frame coded MBs.

20 Although the invention has been described in connection with various specific embodiments, those skilled in the art will appreciate that numerous adaptations and modifications may be made thereto without departing from the spirit and scope of the 25 invention as set forth in the claims.

4.Brief Description of Drawings

FIG. 1 is an illustration of a video object plane (VOP) coding and decoding process in accordance with the present invention.

5 FIG. 2 is a block diagram of an encoder in accordance with the present invention.

FIG. 3 illustrates an interpolation scheme for a half-pixel search.

10 FIG. 4 illustrates direct mode coding of the top field of an interlaced-coded B-VOP in accordance with the present invention.

FIG. 5 illustrates direct mode coding of the bottom field of an interlaced-coded B-VOP in accordance with the present invention.

15 FIG. 6 illustrates reordering of pixel lines in an adaptive frame/field prediction scheme in accordance with the present invention.

FIG. 7 is a block diagram of a decoder in accordance with the present invention.

20 FIG. 8 illustrates a macroblock layer structure in accordance with the present invention.

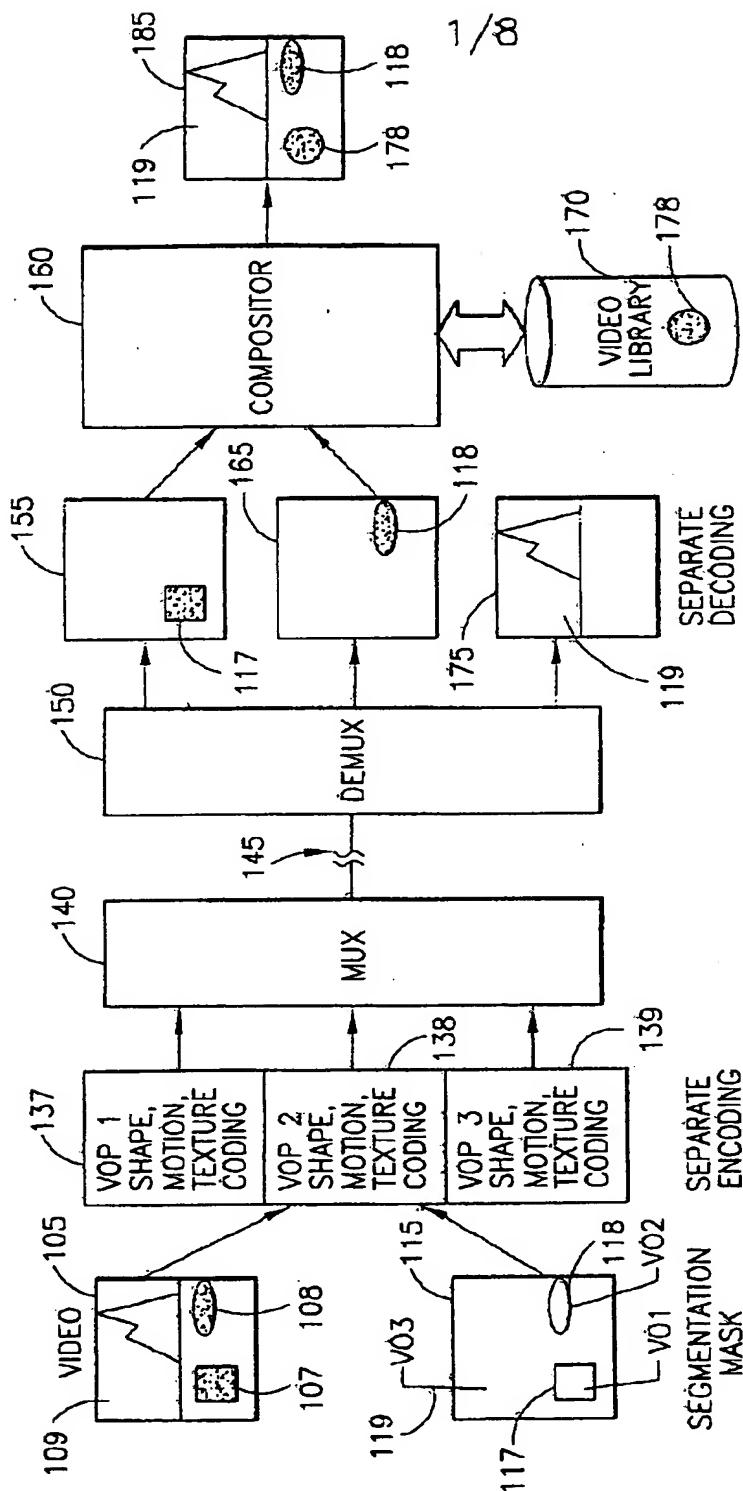


FIG. 1

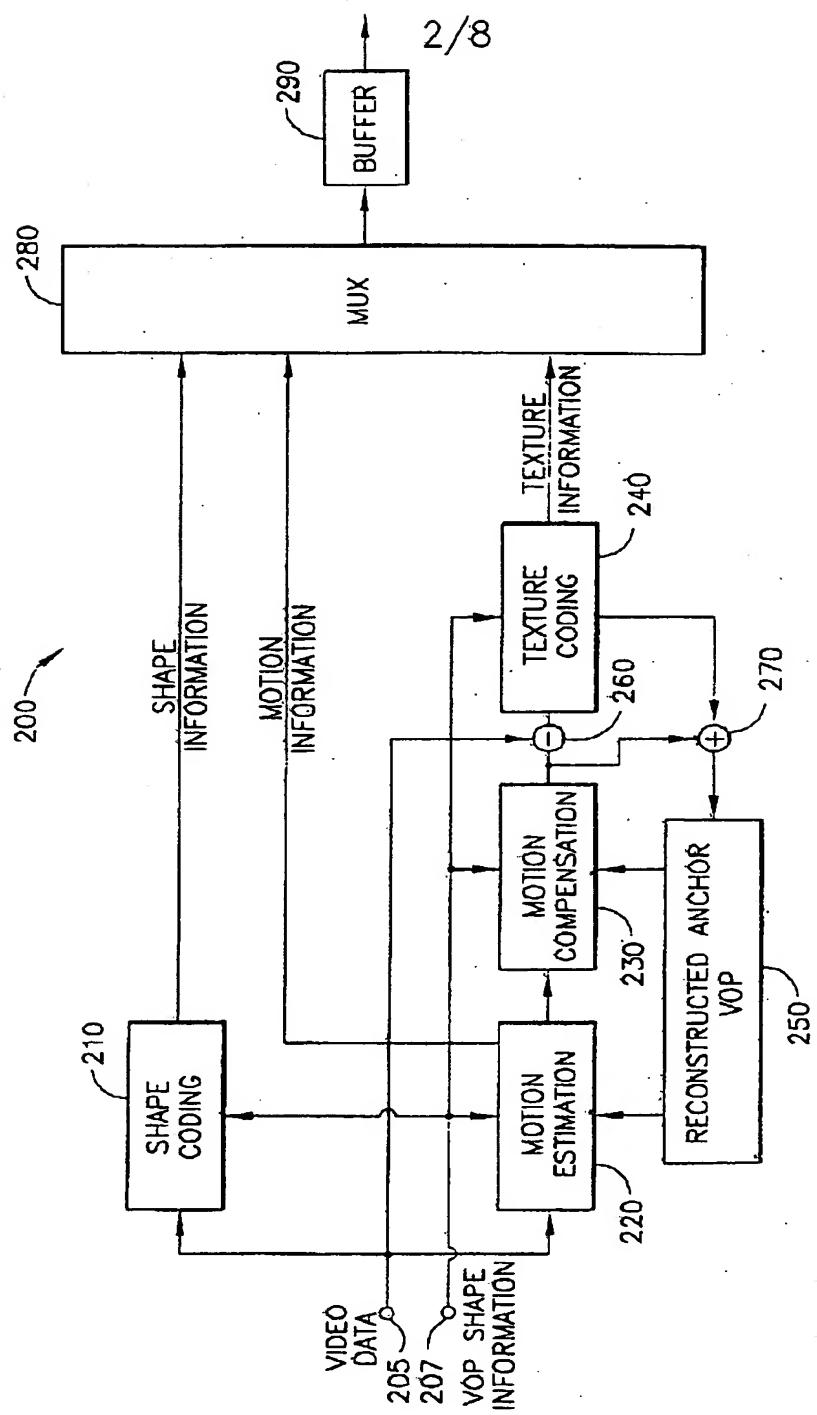


FIG.2

3/8

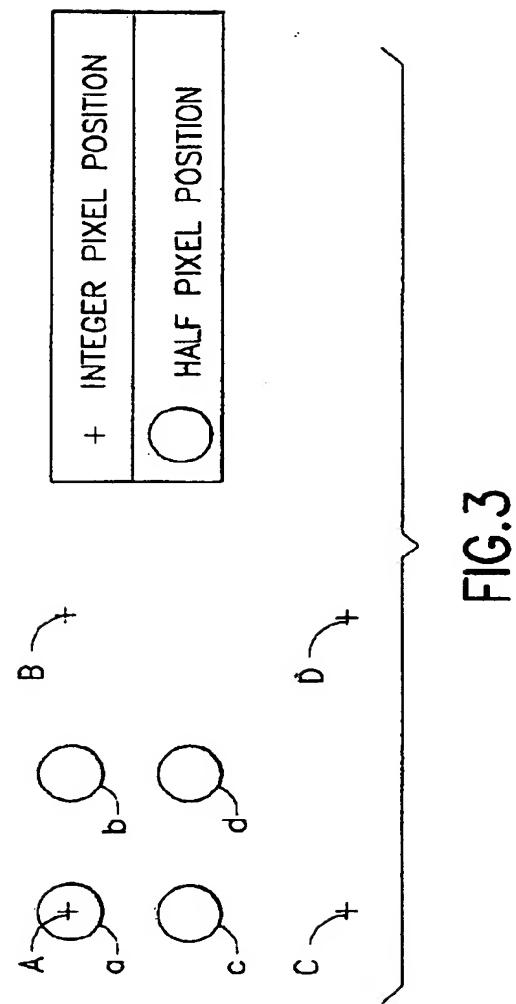


FIG.3

4/8

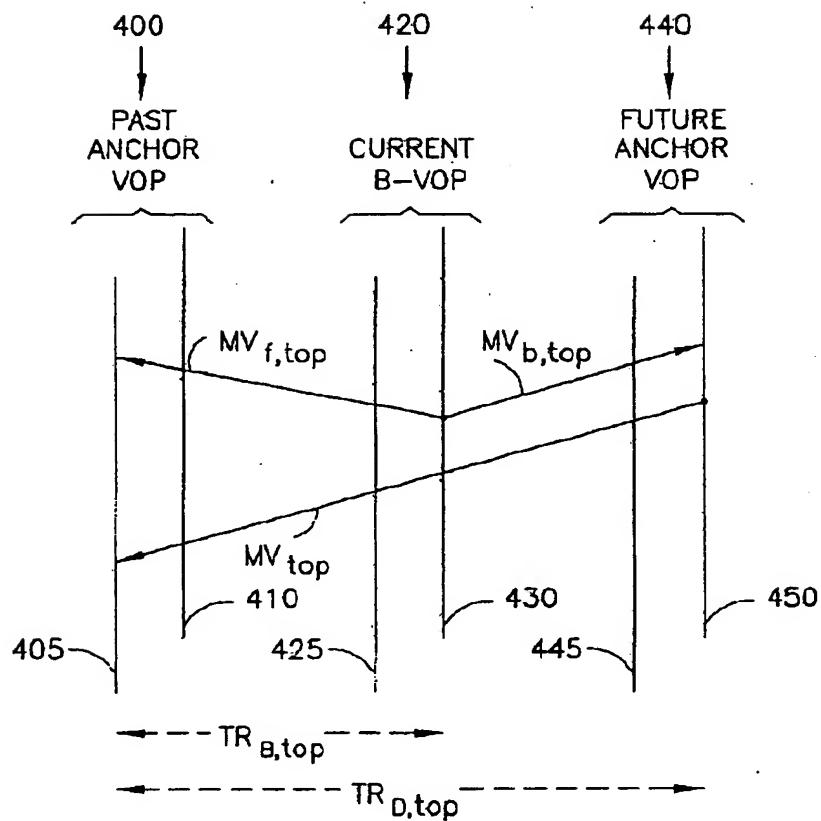


FIG.4

5/8

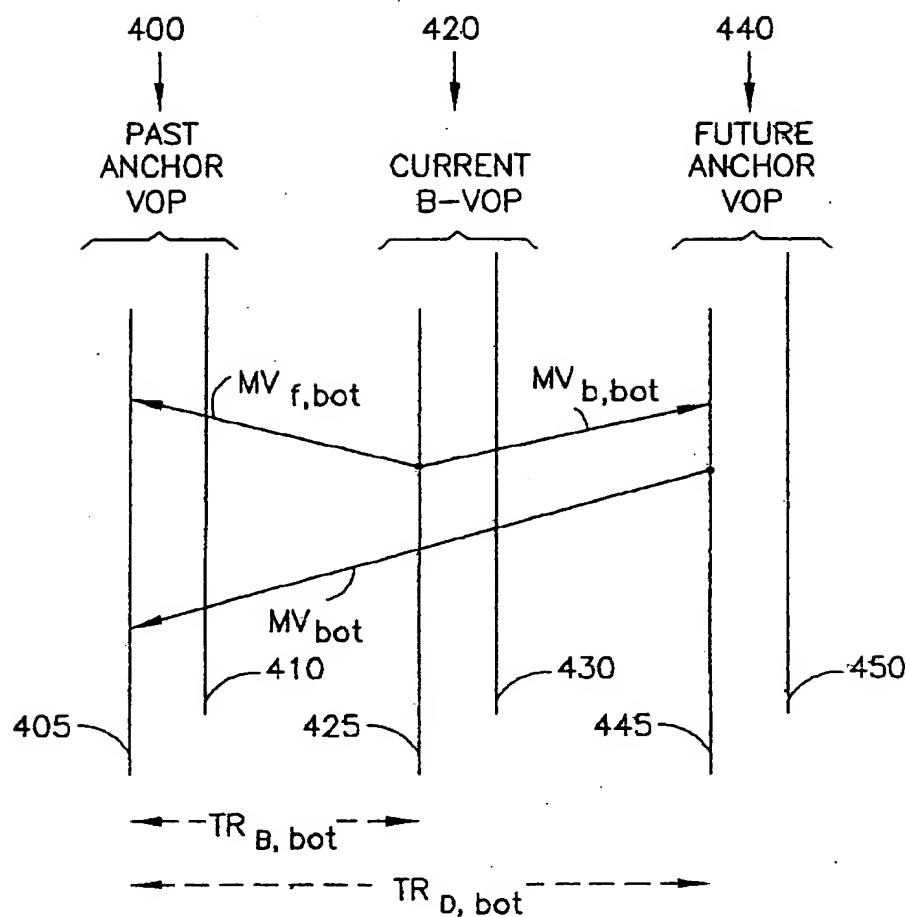


FIG.5

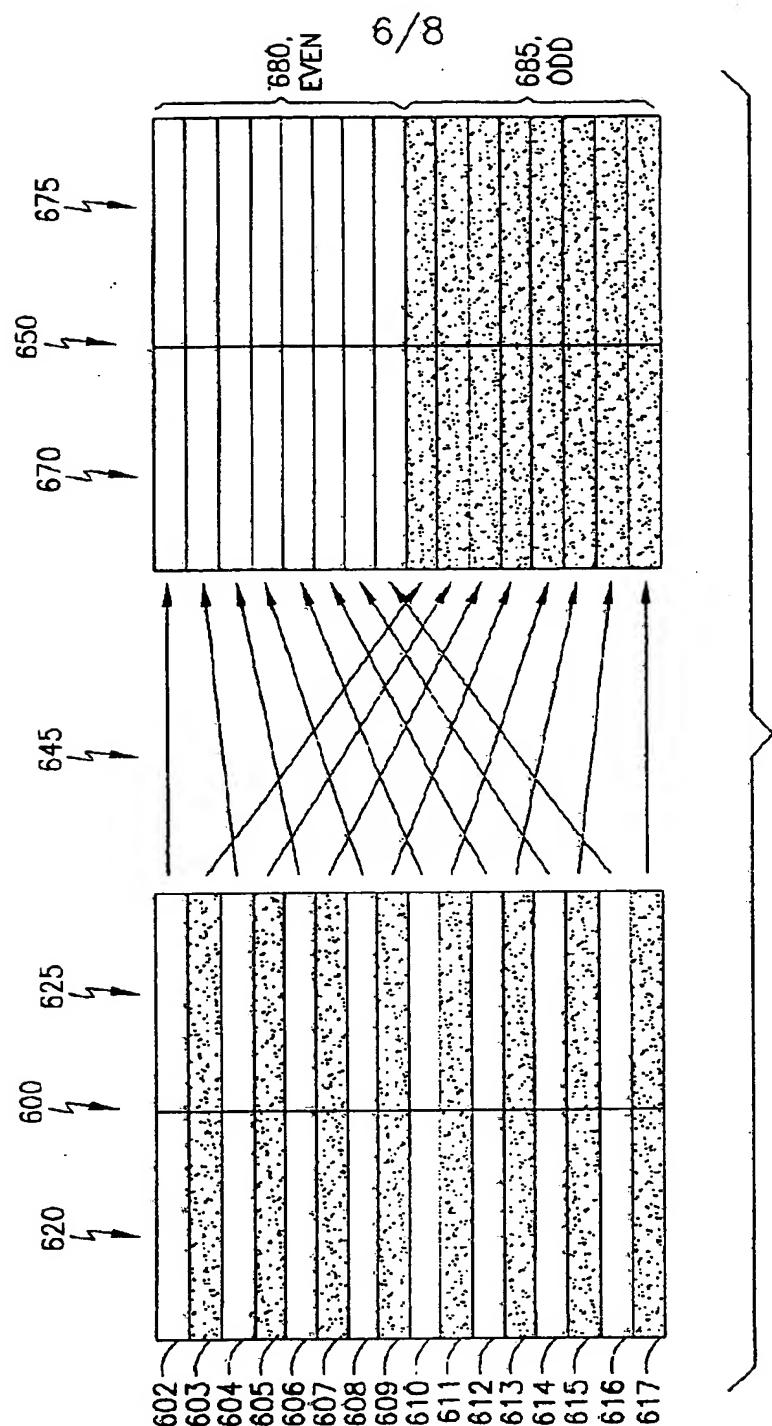


FIG.6

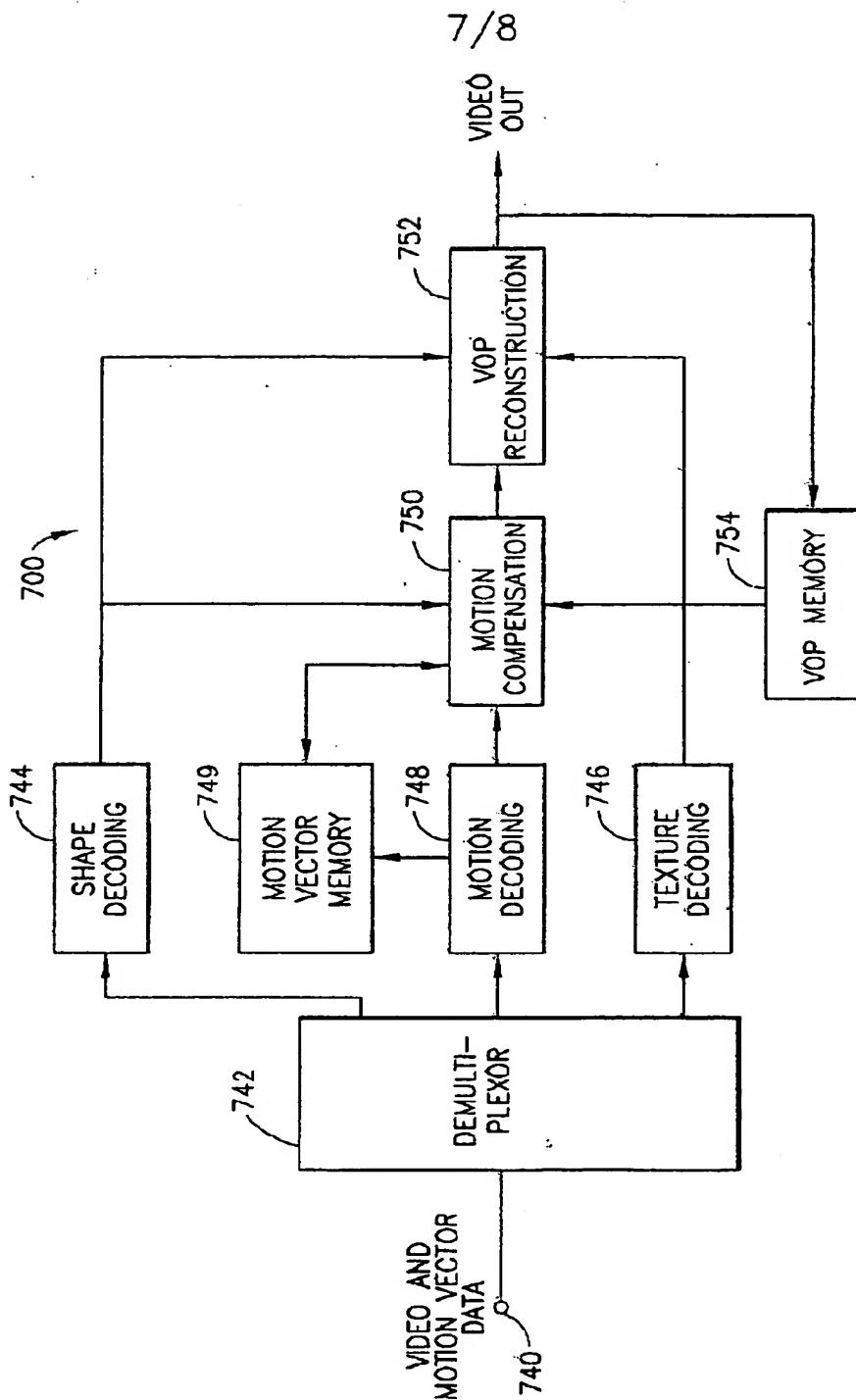


FIG.7

8/8

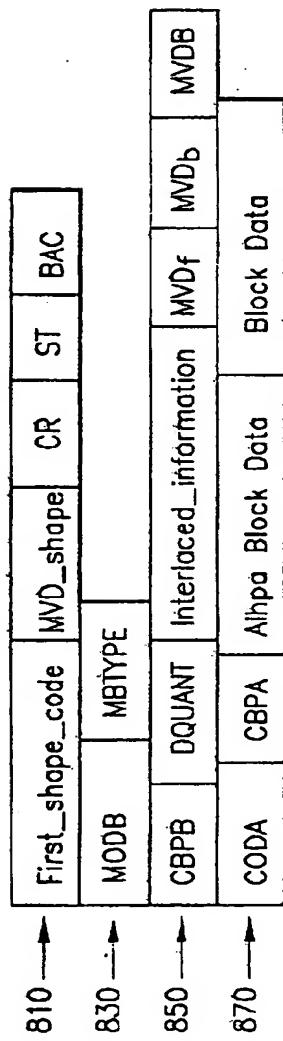


FIG.8

1. Abstract

A system for coding of digital video images such as bi-directionally predicted video object planes (B-VOPs) (420), in particular, where the B-VOP and/or a reference image (400,440) used to code the B-VOP is interlaced coded. For a B-VOP macroblock (420) which is co-sited with a field predicted macroblock of a future anchor picture (440), direct mode prediction is made by calculating four field motion vectors ($MV_{f,top}$, $MV_{f,bot}$, $MV_{b,top}$, $MV_{b,bot}$), then generating the prediction macroblock. The four field motion vectors and their reference fields are determined from (1) an offset term (MV_b) of the current macroblock's coding vector, (2) the two future anchor picture field motion vectors (MV_{top} , MV_{bot}), (3) the reference field (405,410) used by the two field motion vectors of the co-sited future anchor macroblock, and (4) the temporal spacing ($TR_{b,top}$, $TR_{b,bot}$, $TR_{f,top}$, $TR_{f,bot}$), in field periods, between the current B-VOP fields and the anchor fields. Additionally, a coding mode decision process for the current MB selects a forward, backward, or average field coding mode according to a minimum sum of absolute differences (SAD) error which is obtained over the top (430) and bottom (425) fields of the current MB (420).

2. Representative Drawing

Figure 1

54

**This Page is Inserted by IFW Indexing and Scanning
Operations and is not part of the Official Record**

BEST AVAILABLE IMAGES

Defective images within this document are accurate representations of the original documents submitted by the applicant.

Defects in the images include but are not limited to the items checked:

- BLACK BORDERS**
- IMAGE CUT OFF AT TOP, BOTTOM OR SIDES**
- FADED TEXT OR DRAWING**
- BLURRED OR ILLEGIBLE TEXT OR DRAWING**
- SKEWED/SLANTED IMAGES**
- COLOR OR BLACK AND WHITE PHOTOGRAPHS**
- GRAY SCALE DOCUMENTS**
- LINES OR MARKS ON ORIGINAL DOCUMENT**
- REFERENCE(S) OR EXHIBIT(S) SUBMITTED ARE POOR QUALITY**
- OTHER:** _____

IMAGES ARE BEST AVAILABLE COPY.

As rescanning these documents will not correct the image problems checked, please do not report these problems to the IFW Image Problem Mailbox.